sistemas de comunicación

A. Bruce Carlson

SISTEMAS DE COMUNICACIÓN

A. BRUCE CARLSON

Profesor Adjunto de Ingeniería de Sistemas,
Rensselaer Polytechnic Institute

Traductor:

JOSÉ REFUGIO SALAS CONTRERAS
Mayor Técnico en Transmisiones Militares,
Catedrático de la Escuela Militar de Transmisiones,
Ex Catedrático de la Facultad de Ingeniería
de la Universidad de Guadalajara

Revisor Técnico:

FRANCISCO HERNÁNDEZ RANGEL
Profesor de Comunicaciones de la
Escuela Superior de Ingeniería Mecánica
y Eléctrica del Instituto Politécnico Nacional.
Profesor de Teoría Electromagnética
de la Facultad de Ingeniería de la
Universidad Nacional Autónoma

McGRAW-HILL

MÉXICO • BOGOTÁ • BUENOS AIRES • CARACAS • GUATEMALA • LISBOA
MADRID • NUEVA YORK • PANAMÁ • SAN JUAN • SANTIAGO • SÃO PAULO
AUCKLAND • HAMBURGO • LONDRES • MONTREAL • NUEVA DELHI
PARÍS • SAN FRANCISCO • SINGAPUR • ST. LOUIS
SIDNEY • TOKIO • TORONTO

CONTENIDO

PREFACIO

Este texto y su primera edición son una introducción a los sistemas de comunicación eléctrica, ambos escritos a un nivel apropiado para estudiantes a punto de graduarse o que están en su primer año de graduados. Como la comunicación eléctrica encierra diversos aspectos que van desde las arduas consideraciones electrónicas y electromagnéticas hasta las abstracciones matemáticas inherentes a la teoría de la información y de la detección, he intentado trazar aquí una trayectoria media, tomando los punto de vista de la ingeniería de sistemas enfocados sobre los conceptos básicos, técnicas y problemas que caracterizan a la transferencia de la información por medio de señales eléctricas, dadas las inevitables limitaciones de los sistemas físicos. El tema específico y su organización está expuesto en el contenido y se trata en la Sec. 1.6. Los motivos de esta nueva edición son los siguientes:

1. Se introducen por primera vez los fundamentos de la transmisión de señales analógicas y digitales en el contexto de la comunicación en banda base (Cap. 4). Así, conceptos como pérdidas en canales, distorsión, requisitos de ancho de banda, relaciones señal a ruido y errores, son presentados a un nivel más elemental y se establece una norma de comparación para el estudio subsecuente de los métodos de modulación.

2. En atención a la importancia creciente de la comunicación digital, la cobertura de este tema se ha ampliado e intercalado a lo largo del texto, comenzando en el capítulo 4.

3. Mayor flexibilidad tanto para el instructor como para el estudiante, pues ratamos en capítulos separados temas como: ruido en la modulación de onda contiua (Cap. 7), muestreo y modulación de pulsos (Cap. 8), teoría de la información Cap. 9) y sistemas digitales de datos (Cap. 10), más tres Apéndices. Se tiene así, un ran campo para la selección de temas a partir del Cap. 6. Por otra parte, el material pcional de naturaleza más avanzada, en cada capítulo, ha sido señalado con el ímbolo *.

4. Las funciones de correlación y densidad espectral han sido incluidas priero para señales determinísticas (Sec. 2.6) usando el concepto general de productos scalares, por medio del cual se allana el camino para el análisis de las señales aleatoias (Sec. 3.5). La teoría espacial de la señal que se sustenta en los productos escalaes, se trata en el Apéndice A.

5. Hay muchos ejemplos, escogidos particularmente para ilustrar las técnicas n que la mayoría de los estudiantes encuentran dificultades. Hay también ejercicios generalmente con respuestas) diseminados en el texto para motivarlos a evaluar su omprensión del material.

6. Se desarrolló un esfuerzo especial para servir a las necesidades de dos gruos de estudiantes: para los que están comenzando una secuencia de estudio en el rea de la comunicación-eléctrica; y para los que desean que se trate esa materia con elativo detenimiento. Para los primeros, el material opcional constituye un enlace ara el trabajo a nivel de graduado. Para los segundos, he señalado las relaciones ntre campos relacionados con otras disciplinas, y la ingeniería eléctrica.

De conformidad con la heterogeneidad de intereses y un deseo de mantener los rerrequisitos a un mínimo, he supuesto que los estudiantes son moderadamente deptos al análisis de circuitos (incluyendo problemas simples de transitorios y estao estacionario) y poseen algunos conocimientos de electrónica. A propósito de ello iré que los estudios previos sobre análisis de sistemas lineales, transformadas de ourier y teoría de la probabilidad son útiles pero no esenciales. Bajo esta base, los rimeros ocho o nueve capítulos pueden ser estudiados en un semestre si se omiten s secciones opcionales. Con una selección de las partes que se agregarán u omitin, se armará un texto adecuado a cursos de variables duraciones o para estudiantes on diferentes bases.

Han sido incluidos dos tipos de referencias: las que se citan por medio de notas pie de página y dan mayores detalles de un párrafo o artículo, y las mencionadas jo el encabezado de Lecturas Suplementarias Selectas que constituyen una breve lación bibliográfica de algunos tratados más elaborados. Todas estas referencias n sido escogidas por su fácil localización.

También se han incluido aproximadamente 370 problemas, contados los que guran en los apéndices. Ellos abarcan desde cálculos simples (con soluciones cluidas) hasta análisis más avanzados y problemas de diseño. Los problemas que refieren al material opcional o que requieren cierta ingeniosidad del estudiante, tán identificados con el símbolo *. Esto puede ser considerado por el instructor mo tema de discusión en clase. Si se solicita al editor un manual del instructor que ntiene soluciones completas a los problemas, él podrá proporcionarlo.

Además de soluciones a problemas, el manual del instructor contiene: Figuras seleccionadas del texto en forma adecuada para hacer transparencias proyectables, instrucciones para varias demostraciones de lectura y una completa guía de estudios para aquellos que deseen enseñar la materia por medio de la instrucción individualizada o autodidacta. Casualmente, he usado este tipo de instrucción en Rensselaer desde 1972, por ello esas guías de estudio están basadas en una considerable experiencia. El método parece funcionar muy bien con esta materia y la respuesta del estudiante ha sido generalmente entusiasta.

Estoy en deuda con muchas personas por sus consejos y auxilio para esta segunda edición. El Dr. David G. Gisser, del Instituto Politécnico de Rensselaer, aportó importantes contribuciones relacionadas con la electrónica de la comunicación. Otros colegas de Rensselaer que me auxiliaron son los doctores Dean N. Arden, Charles M. Close, Lester A. Gerhardt y Dean K. Frederick. El Dr. John C. Lindenlaub, de la Universidad de Purdue, proporcionó muchas sugerencias importantes para que fuera realizada la segunda edición, y colaboró en las guías de estudio para estudiantes, las cuales han sido probadas en clase, en Purdue. Se hicieron numerosas mejoras pedagógicas con los comentarios de estudiantes y auxiliares de enseñanza graduados en el curso de instrucción autodidacta, principalmente los señores Mounir P. Badawy, Nelson R. Corby, Jeffrey S. Lucash, Michael P. Meyer y Arthur P. Sarkisian. Finalmente, agradezco el apoyo que recibí durante este trabajo a la administración de la Escuela de Ingeniería en Rensselaer, a mi esposa y a mi familia.

A. BRUCE CARLSON

1

INTRODUCCIÓN

*Attention, the Universe! By kingdoms, right wheel!** Esta frase profética es el primer mensaje telegráfico registrado, que envió a través de una línea de 16 km Samuel F. B. Morse, en 1838. Así nació una nueva era en la comunicación, la era de la comunicación eléctrica.

Hoy, los sistemas de comunicación eléctrica se encuentran en todas partes donde se transmite información de un punto a otro. El teléfono, la radio y la televisión han venido a ser factores de la vida diaria. Los circuitos de larga distancia cubren el globo terráqueo llevando texto, voces e imágenes. Los sistemas de radar y telemetría desempeñan papeles importantes, vitales, en navegación, defensa e investigación científica. Las computadoras hablan a otras computadoras por medio de enlaces transcontinentales de datos. Los logros son muchos y la lista es, al parecer, interminable. Ciertamente se han logrado grandes avances desde los días de Morse; también es cierto que en las próximas décadas veremos muchas nuevas hazañas en la ingeniería de la comunicación, aunque las aplicaciones potenciales sólo están limitadas por las necesidades, aspiraciones e imaginación del hombre.

Dados los innumerables tipos de sistemas de comunicación, un libro como este, dirigido a la materia de la comunicación eléctrica, no puede cubrir cada aplicación. El resultado sería un enorme catálogo que pronto estaría anticuado o des-

* "¡Atención, el Universo! Por reinos, ¡a la derecha!" (Voz de mando en el ejército estadounidense.)

actualizado y no podríamos discutir en detalle los componentes de un sistema específico. Un sistema típico está formado de numerosas y diversas partes cuya comprensión escapa virtualmente a todas las especialidades de la ingeniería eléctrica: conversión de la energía, teoría de redes, electrónica y teoría electromagnética, para citar unas cuantas. Más aún, en un análisis pormenorizado se perdería el punto esencial de que un sistema es un todo integrado que realmente excede la suma de sus partes.

En vez de eso, este libro trata el tema desde un punto de vista más general. Reconociendo que todos los sistemas de comunicación tienen la misma función básica, o sea, *la transmisión de la información*, buscaremos y aislaremos los principios y problemas de la información que se ha de transmitir en forma eléctrica. Estos serán examinados detalladamente y se harán los análisis fundamentales y las técnicas de diseño aplicables a cualquier tipo de comunicación eléctrica. En pocas palabras, este libro trata a los sistemas de comunicación como *sistemas*.

El contacto con la ingeniería de sistemas es algo poderoso, y al mismo tiempo algo abstracto, pero confiando en las matemáticas, podemos llegar al corazón de los más complejos problemas. Sin embargo, debemos tener en mente que las matemáticas son un medio para lograr un fin, y éste sigue siendo un conocimiento básico de la comunicación eléctrica. El uso de la herramienta matemática abstracta y sus modelos deben ser dirigidos a ese fin, con razonamiento físico y juicio de ingeniería.

El propósito de este capítulo de introducción es doble: dar una descripción general de los sistemas de comunicación, enumerar los elementos esenciales, sus funciones y problemas asociados, y situar en perspectiva los papeles que desempeñan las disciplinas en la ingeniería de los sistemas de la comunicación.

1.1 COMUNICACIÓN, MENSAJES Y SEÑALES

Para empezar, definimos a la comunicación como el proceso por medio del cual la información se transfiere de un punto llamado la *fuente*, en espacio y tiempo, a otro punto que es el *destino* o usuario. Un sistema de comunicación es la totalidad de mecanismos que proporcionan el enlace para la información entre fuente y destino. Un sistema de comunicación eléctrica es aquel que ejecuta esta función principal, pero no exclusivamente, por medio de dispositivos y fenómenos eléctricos.

Es evidente que el concepto de *información* es la esencia de la comunicación. Pero, ¿qué es la información? Aquí podemos entrar en dificultades, porque información es una palabra embarazosa que implica nociones semánticas y filosóficas tales como conocimiento o significado. Estos aspectos subjetivos, aunque pueden ser fascinantes, no son aplicables a los problemas tecnológicos de la comunicación. Eventualmente, en nuestra discusión de la teoría de la información, se formulará una definición matemática precisa de información. Por lo pronto, sin embargo, se seguirá una ruta ligeramente diferente. En vez de abordar la información por sí misma, nos concentraremos en la manifestación física de la información como es producida por la fuente, es decir, el mensaje.

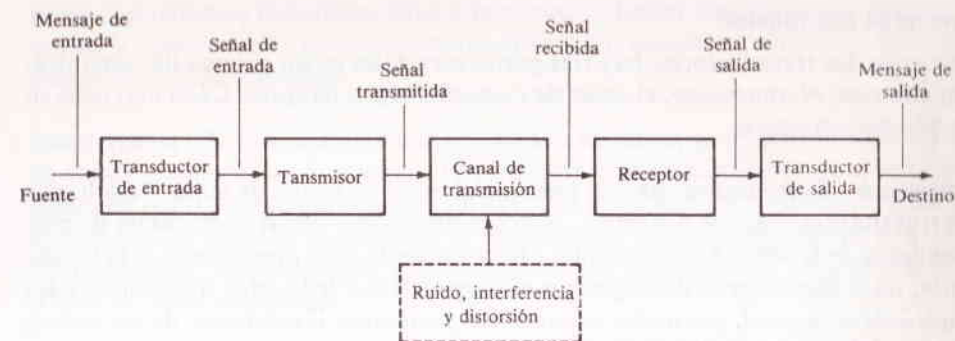



FIGURA **1.1**
Los elementos de un sistema de comunicación.

Hay muchas clases de fuentes de información, incluso hombres y máquinas; por eso, los mensajes aparecen en muchas formas: una secuencia de símbolos o letras discretas (v. gr., palabras escritas en una forma telegráfica, las perforaciones en una tarjeta IBM); una magnitud sencilla variando con el tiempo (v. gr., la presión acústica producida por la voz o la música; la posición angular de un giróscopo); varias funciones del tiempo y otras variables (v. gr., la intensidad de la luz y el color de una escena de televisión). Pero, sea cual fuere el mensaje, el objeto de un sistema de comunicación es proporcionar una réplica aceptable de él en su destino.

Como regla, establezcamos que el mensaje producido por una fuente no es eléctrico y, por lo tanto, es necesario un transductor de entrada. Este transductor convierte el mensaje en una *señal*, una magnitud eléctrica variable, tal como un voltaje o una corriente. Similarmente, otro transductor en el destino convierte la señal de salida a la forma apropiada del mensaje. Si bien el diseño de los transductores es una parte importante de la ingeniería de la comunicación, limitaremos nuestra consideración a la parte estrictamente eléctrica del sistema, es decir, aquella porción donde el mensaje aparece como una señal eléctrica. En lo sucesivo, los términos *señal* y *mensaje* se usarán indistintamente, puesto que tanto la señal como el mensaje son una materialización física de la información.

Establecidos estos preliminares, volvamos nuestra atención al sistema, sus partes y sus problemas.

1.2 LOS ELEMENTOS DE UN SISTEMA DE COMUNICACIÓN

La figura 1.1 muestra los elementos funcionales de un sistema completo de comunicación. Por conveniencia, los hemos aislado como entidades distintas, aunque en los sistemas reales la separación no puede ser tan obvia. También se indica que hay algunos factores no deseados que inevitablemente forman parte de la figura.

Elementos funcionales

Omitiendo los transductores, hay tres partes esenciales en un sistema de comunicación eléctrica, *el transmisor*, el *canal de transmisión* y el *receptor*. Cada uno tiene su función característica.

Transmisor El transmisor pasa el mensaje al canal en forma de señal. Para lograr una transmisión eficiente y efectiva, se deben desarrollar varias operaciones de procesamiento de la señal. La más común e importante de estas operaciones es la *modulación*, un proceso que se distingue por el acoplamiento de la señal transmitida a las propiedades del canal, por medio de una onda portadora. Hablaremos de este proceso más adelante, en cuatro capítulos.

Canal de transmisión El canal de transmisión o medio es el enlace eléctrico entre el transmisor y el receptor, siendo el puente de unión entre la fuente y el destino. Puede ser un par de alambres, un cable coaxial, una onda de radio o un rayo laser. Pero sin importar el tipo, todos los medios de transmisión eléctricos se caracterizan por la *atenuación*, la disminución progresiva de la potencia de la señal conforme aumenta la distancia. La magnitud de la atenuación puede ser pequeña o muy grande. Generalmente es grande y, por lo tanto, es un factor que debe ser considerado.

Receptor La función del receptor es extraer del canal la señal deseada y entregarla al transductor de salida. Como las señales son frecuentemente muy débiles, como resultado de la atenuación, el receptor debe tener varias etapas de amplificación. En todo caso, la operación clave que ejecuta el receptor es la *demodulación* (o detección), el caso inverso del proceso de modulación del transmisor, con lo cual vuelve la señal a su forma original.

Contaminaciones

Durante la transmisión de la señal ocurren ciertos efectos no deseados. Uno de ellos es la atenuación, la cual reduce la *intensidad* de la señal; sin embargo, son más serios la distorsión, la interferencia y el ruido, los cuales se manifiestan como alteraciones de la *forma* de la señal. Al introducirse estas contaminaciones al sistema, es una práctica común y conveniente imputárselas al canal, pues el transmisor y el receptor son considerados ideales. En la figura 1.1 se ilustra esta conversión.

En términos generales, cualquier perturbación no intencional de la señal se puede clasificar como ''ruido'', y algunas veces es difícil distinguir las diferentes causas que originan una señal contaminada. Existen buenas razones y bases para separar esos tres efectos, de la manera siguiente:

Distorsión Es la alteración de la señal debida a la respuesta imperfecta del sistema a ella misma. A diferencia del ruido y la interferencia, la distorsión desaparece cuando la señal deja de aplicarse. El diseño de sistemas perfeccionados o redes de compensación reduce la distorsión. En teoría es posible lograr una compensación perfecta. En la práctica debe permitirse cierta distorsión, aunque su magnitud debe estar dentro de límites tolerables.

Interferencia Es la contaminación por señales extrañas, generalmente artificiales y de forma similar a las de la señal. El problema es particularmente común en emisiones de radio, donde pueden ser captadas dos o más señales simultáneamente por el receptor. La solución al problema de interferencia es obvia: eliminar en una u otra forma la señal interferente o su fuente. En este caso es posible una solución perfecta, si bien no siempre práctica.

Ruido Finalmente, dejando lo peor al último, llegamos al ruido. Por ruido se debe entender las señales aleatorias e impredecibles de tipo eléctrico originadas en forma natural dentro o fuera del sistema. Cuando estas variaciones se agregan a la señal portadora de la información, ésta puede quedar en gran parte oculta o eliminada totalmente. Por supuesto que podemos decir lo mismo en relación a la interferencia y la distorsión y en cuanto al ruido que no puede ser eliminado nunca completamente, ni aun en teoría. Como veremos, el ruido no eliminable es uno de los problemas básicos de la comunicación eléctrica.

1.3 MODULACIÓN

Muchas señales de entrada no pueden ser enviadas directamente hacia el canal, como vienen del transductor. Para eso se modifica una onda portadora, cuyas propiedades se adaptan mejor al medio de transmisión en cuestión, para representar el mensaje. La modulación es la alteración sistemática de una onda portadora de acuerdo con el mensaje (señal moduladora) y puede ser también una codificación.

Es interesante hacer hincapié en que muchas formas de comunicación no eléctricas también encierran un proceso de modulación, y la voz es un buen ejemplo. Cuando una persona habla, los movimientos de la boca ocurren de una manera más bien lenta, del orden de los 10 Hz†, que realmente no pueden producir ondas acústicas que se propaguen. La transmisión de la voz se hace por medio de la generación de tonos portadores, de alta frecuencia, en las cuerdas vocales, tonos que son modulados por los músculos y órganos de cavidad oral. Lo que el oído capta como voz, es una onda acústica modulada, muy similar a una onda eléctrica modulada.

Tipos de modulación

El éxito de un sistema de comunicación en una misión determinada, depende en gran parte de la modulación, tan es así que el tipo de modulación es una decisión alrededor de la cual gravita el diseño del sistema y por esta razón muchas técnicas de mo-

† El *hertz*, abreviado Hz, es la unidad de frecuencia en ciclos por segundo. Se usará a menudo en este libro con prefijos, como: kHz, MHz

dulación han evolucionado y cubierto diversas tareas y requisitos de muchos sistemas. Y conforme aparezcan nuevas exigencias, se desarrollarán nuevas técnicas.

A pesar de la multitud de variedades, es posible identificar dos tipos básicos de modulación en relación a la clase de onda portadora: la *modulación de onda continua* (CW), en la cual la portadora es simplemente una forma de onda senoidal, y la *modulación de pulsos*, en la cual la portadora es un tren periódico de pulsos.

Puesto que la modulación de onda continua es un proceso continuo, es posible adaptarla a señales que están variando constantemente con el tiempo. Por lo general, la portadora senoidal es de mayor frecuencia que cualquiera de las componentes de frecuencia contenidas en la señal moduladora. El proceso de modulación se caracteriza pues por una *traslación de frecuencia*, es decir, el espectro del mensaje (su contenido de frecuencia) se corre hacia arriba a otra banda de mayor frecuencia.

La Modulación de pulsos es un proceso discontinuo o discreto, en el sentido de que los pulsos aparecen sólo en ciertos intervalos de tiempo. Por eso la modulación de pulsos se adapta mejor a los mensajes que son discretos por naturaleza. Con la ayuda del *muestreo*, las señales que varían continuamente pueden ser transmitidas sobre portadoras pulsadas. A menudo, tanto en los telégrafos como en los teletipos, la modulación de pulsos y la codificación van de la mano.

Como alternativa a la clasificación anterior, algunas veces es preferible designar a la modulación como *analógica* o *codificada* (digital). Esto es particularmente cierto en los sistemas más complejos que emplean ambas técnicas (modulación de CW y pulsada), haciendo distinción de su tipo indefinido de portadora. La diferencia entre análogica y digital es la siguiente: en la modulación analógica, el parámetro modulado varía en razón directa a la señal moduladora. En la modulación codificada, ocurre una transformación digital, por medio de la cual el mensaje se cambia de un lenguaje simbólico a otro. Si el mensaje es originalmente una función continua del tiempo, debe ser muestreado y digitalizado (cuantificado) antes de ser codificado.

Pero haciendo caso omiso del tipo —CW o pulsada, analógica o codificada— la modulación debe ser un proceso *reversible*, de tal manera que el mensaje pueda ser recuperado en el receptor por medio de la operación complementaria de demodulación.

¿Para qué se modula?

En su libro sobre la teoría de la modulación, Black (1953)† dedica un capítulo a esta interrogante. Hemos dado ya una respuesta concisa: la modulación se precisa para acoplar la señal con el medio de transmisión. Sin embargo, este acoplamiento implica varias consideraciones que merecen una ulterior ampliación.

Modulación por facilidad de radiación Una radiación eficiente de energía electromagnética requiere de elementos radiadores (antenas) cuyas dimensiones físicas sean

† Las referencias se indican en esta modalidad a lo largo del texto. Las citas completas están enumeradas por el autor en las Referencias, al final del libro.

por lo menos de 1/10 de su longitud. Pero muchas señales, especialmente de audio, tienen componentes de frecuencia del orden de los 100 Hz o menores, para lo cual necesitan antenas de unos 300 km de longitud si se radiaran directamente. Utilizando la propiedad de traslación de frecuencia de la modulación, estas señales se pueden imprimir sobre una portadora de alta frecuencia, con lo que se logra una reducción sustancial del tamaño de la antena. Por ejemplo, en la banda de radio de FM, donde las portadoras están en el intervalo de 88 a 108 MHz, las antenas no deben ser mayores de un metro.

Modulación para reducir el ruido y la interferencia Hemos dicho que es imposible eliminar totalmente el ruido del sistema. Y aunque es posible eliminar la interferencia, puede no ser práctico. Por fortuna, ciertos tipos de modulación tienen la útil propiedad de suprimir tanto el ruido como la interferencia. La supresión, sin embargo, ocurre a un cierto precio; generalmente requiere de un ancho de banda de transmisión (intervalo de frecuencia) mucho mayor que el de la señal original; de ahí la designación de *reducción del ruido de banda ancha*. Este convenio de ancho de banda para la reducción del ruido es uno de los interesantes y a veces desventajosos aspectos del diseño de un sistema de comunicación.

Modulación por asignación de frecuencia El propietario de un aparato de radio o televisión puede seleccionar una de varias estaciones, aun cuando todas las estaciones estén transmitiendo material de programa similar en el mismo medio de transmisión. Es posible seleccionar y separar cualquiera de las estaciones, dado que cada una tiene asignada una frecuencia portadora diferente. Si no fuera por la modulación, sólo operaría una estación en un área dada. Dos o más estaciones que transmitan directamente en el mismo medio, sin modulación, producirían una mezcla inútil de señales interferentes.

Modulación para multicanalización A menudo se desea transmitir muchas señales en forma simultánea entre dos puntos. Las técnicas de multicanalización son formas, intrínsecas de modulación, permiten la transmisión de señales múltiples sobre un canal, de tal manera que cada señal puede ser captada en el extremo receptor. Las aplicaciones de la multicanalización comprenden telemetría de datos, emisión de FM estereofónica y telefonía de larga distancia. Es muy común, por ejemplo, tener hasta 1 800 conversaciones telefónicas de ciudad a ciudad, multicanalizadas y transmitidas sobre un cable coaxial de un diámetro menor de un centímetro.

Modulación para superar las limitaciones del equipo El diseño de un sistema queda generalmente a la disponibilidad de equipo, el cual a menudo presenta inconvenientes en relación con las frecuencias involucradas. La modulación se puede usar para situar una señal en la parte del espectro de frecuencia donde las limitaciones del equipo sean mínimas o donde se encuentren más fácilmente los requisitos de diseño. Para este propósito, los dispositivos de modulación se encuentran también en los receptores, como ocurre con los transmisores.

1.4 LIMITACIONES FUNDAMENTALES EN LA COMUNICACIÓN ELÉCTRICA

En el diseño de un sistema de comunicación o de cualquier sistema para esta materia, el ingeniero se coloca frente a dos clases generales de restricciones: por un lado, los *factores tecnológicos*, es decir, los factores vitales de la ingeniería y por otra parte, las *limitaciones físicas fundamentales* impuestas por el propio sistema, o sean, las leyes de la naturaleza en relación con el objetivo propuesto.

Puesto que la ingeniería es, o debe ser, el arte de lo posible, ambas clases de restricciones deben ser analizadas al diseñar el sistema. Hay más de una diferencia, pues los problemas tecnológicos son problemas de practicabilidad que incluyen consideraciones tan diversas como disponibilidad del equipo, interacción con sistemas existentes, factores económicos, etc., problemas que pueden ser resueltos en teoría, aunque no siempre de una manera práctica. Pero las limitaciones físicas fundamentales son justamente eso; cuando aparecen en primer plano, no existen recursos, incluso en teoría. No obstante, los problemas tecnológicos son las limitaciones que en última instancia señalan si pueden o no ser salvadas. Las limitaciones fundamentales en la transmisión de la información por medios eléctricos son *el ancho de banda* y *el ruido*.

La limitación del ancho de banda

Aunque no en forma explícita, la figura 1.1 muestra el elemento tiempo como una parte integrante de los sistemas de comunicación. La utilización de sistemas eficientes conduce a una reducción del tiempo de transmisión, es decir, que se transmite una mayor información en el menor tiempo. Una transmisión de información rápida se logra empleando señales que varían rápidamente con el tiempo. Pero estamos tratando con un sistema eléctrico, el cual cuenta con energía almacenada; y hay una ley física bien conocida que expresa que en todos los sistemas, excepto en los que no hay pérdidas, un cambio en la energía almacenada requiere una cantidad definida de tiempo. Así, no podemos incrementar la velocidad de la señalización en forma arbitraria, ya que en consecuencia el sistema dejará de responder a los cambios de la señal.

Una medida conveniente de la velocidad de la señal es su ancho de banda, o sea, el ancho del espectro de la señal. En forma similar, el régimen al cual puede un sistema cambiar energía almacenada, se refleja en su respuesta de frecuencia útil, medida en términos del *ancho de banda del sistema*. La transmisión de una gran cantidad de información en una pequeña cantidad de tiempo, requiere señales de banda ancha para representar la información y sistemas de banda ancha para acomodar las señales. Por lo tanto, dicho ancho de banda surge como una limitación fundamental. Cuando se requiere de una transmisión en tiempo real, el diseño debe asegurar un adecuado ancho de banda del sistema. Si el ancho de banda es insuficiente, puede ser necesario disminuir la velocidad de señalización, incrementándose así el tiempo de transmisión. A lo largo de estas mismas líneas debe recalcarse que el diseño del equipo no es con mucho un problema de ancho de banda *absoluto* o *fraccionario*, o

sea, el ancho de banda absoluto dividido entre la frecuencia central. Si con una señal de banda ancha se modula una portadora de alta frecuencia, se reduce el ancho de banda fraccional y con ello se simplifica el diseño del equipo. Esta es una razón por qué en señales de TV cuyo ancho de banda es de cerca de 6 MHz se emiten sobre portadoras mucho mayores que en la transmisión de AM, donde el ancho de banda es de aproximadamente 10 Hz.

Asimismo, dado un ancho de banda fraccionario, resultado de las consideraciones del equipo, el ancho de banda absoluto puede incrementarse casi indefinidamente yendo hasta frecuencias portadoras mayores. Un sistema de microondas de 5 GHz puede acomodar 10 000 veces más información en un periodo determinado que una portadora de radiofrecuencia de 500 kHz, mientras que un rayo laser cuya frecuencia sea de 5×10^{14} Hz tiene una capacidad teórica de información que excede al sistema de microondas en un factor de 10^5, o sea, un equivalente aproximado de 10 millones de canales de TV. Por ello es que los ingenieros en comunicaciones están investigando constantemente fuentes de portadoras de alta frecuencia nuevas y utilizables para compensar el factor ancho de banda.

La figura 1.2 muestra aquellas porciones del espectro electromagnético en uso o potencialmente disponible para la comunicación eléctrica. Se indican en ella las aplicaciones representativas y los medios de transmisión. Como guía burda, considérese que el ancho de banda disponible en cualquier punto es aproximadamente el 10% de la frecuencia portadora. Debido a varios factores, el hueco entre los 10^{11} Hz (100 GHz) y los 10^{14} Hz probablemente permanecerá vacío por lo que a comunicación se refiere.

La limitación ruido

Un instrumento de medición que posee un 1% de resolución da lugar a una mayor información que un instrumento con un 10%; la diferencia es 1 de *exactitud*. En forma similar, el éxito en la comunicación eléctrica depende de la exactitud con que el receptor pueda determinar cuál señal es la que fue realmente transmitida, diferenciándola de las señales que podrían haber sido transmitidas. Una identificación perfecta de la señal sería posible sólo en ausencia de ruido y otras contaminaciones, pero el ruido existe siempre en los sistemas eléctricos y sus perturbaciones sobrepuestas limitan nuestra habilidad para identificar correctamente la señal que nos interesa y así, la transmisión de la información.

¿Por qué es inevitable el ruido? Detalle curioso, la respuesta proviene de la teoría cinética. Cualquier partícula a una temperatura diferente de cero absoluto, posee una energía térmica que se manifiesta como movimiento aleatorio o agitación térmica. Si la partícula es un electrón, su movimiento aleatorio origina una corriente aleatoria. Luego, si esta corriente aleatoria ocurre en un medio conductor, se produce un voltaje aleatorio conocido como *ruido térmico* o *ruido de resistencia*. Mientras el ruido de resistencia es sólo una de las posibles fuentes en un sistema, muchos otros están relacionados, en una u otra forma, al movimiento electrónico aleatorio. Más aún, como era de esperarse de la dualidad onda-partícula, existe ruido térmico asociado con la radiación electromagnética. En consecuencia, como no podemos tener

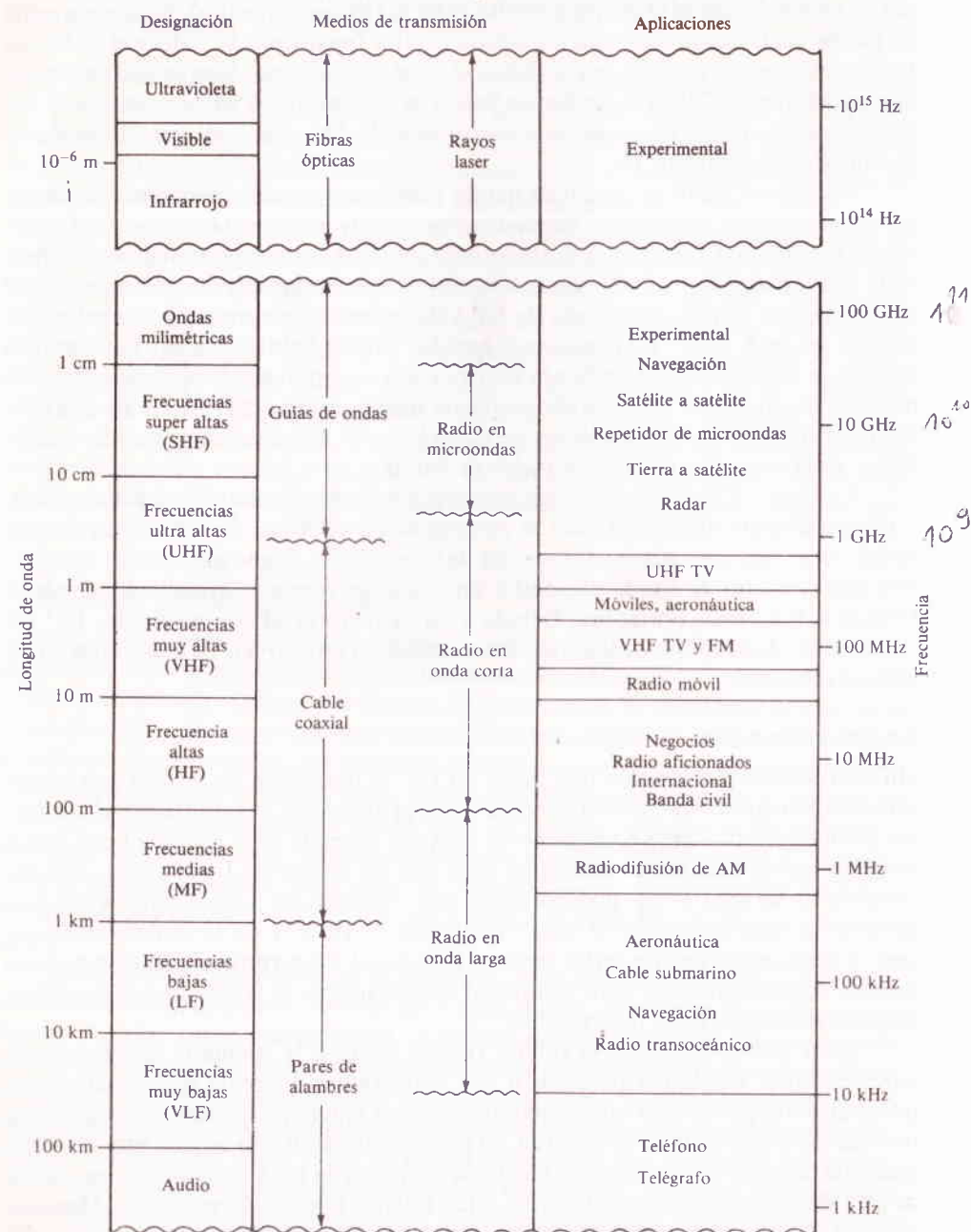



FIGURA 1.2
El espectro electromagnético. (Las aplicaciones mostradas son ilustrativas; las limitaciones de espacio impiden una lista completa).

comunicación eléctrica sin electrones u ondas electromagnéticas, tampoco podemos tener comunicación eléctrica sin ruido.

Las variaciones de ruido típicas son muy pequeñas, del orden de los microvolts. Si las variaciones de la señal son sustancialmente mayores, digamos varios volts pico a pico, el ruido puede ser ignorado. En realidad, en sistemas ordinarios bajo condiciones ordinarias, la relación *señal a ruido* es bastante grande para que el ruido no sea perceptible. Pero en sistemas de amplio régimen o de potencia mínima, la señal recibida puede ser tan pequeña como el ruido o más. Cuando esto suceda la limitación por ruido resulta muy real.

Es importante señalar que si la intensidad de la señal es insuficiente, añadir más pasos de amplificación en el receptor no resuelve nada; el ruido será amplificado junto con la señal, lo cual no mejora la relación señal a ruido. Aumentar la potencia transmitida ayuda, pero la potencia no se puede incrementar en forma indefinida por razón de problemas tecnológicos. (Uno de los primeros cables transatlánticos se deterioró por una ruptura ocasionada por un alto voltaje, aplicado en un esfuerzo por obtener señales útiles en el punto de recepción.) En forma alterna, como se mencionó al principio, podemos permutar el ancho de banda por la relación señal a ruido por medio de técnicas de modulación y codificación. No es de sorprender que la más efectiva de estas técnicas generalmente sea la más costosa y difícil de instrumentar. Nótese también que el trueque del ancho de banda por la relación señal a ruido puede llevarnos de una limitación a otra.

En el análisis final, dado un sistema con ancho de banda y relación señal a ruido fijos, existe un límite superior definido, al cual puede ser transmitida la información por el sistema. Este límite superior se conoce con el nombre de *capacidad* de información y es uno de los conceptos centrales de la teoría de la información. Como la capacidad es finita, se puede decir con apego a la verdad, que el diseño del sistema de comunicación es un asunto de compromiso; un compromiso entre tiempo de transmisión, potencia transmitida, ancho de banda y relación señal a ruido; compromiso de lo más restringido por los problemas tecnológicos.

1.5 CRONOLOGÍA DE LA COMUNICACIÓN ELÉCTRICA

La organización de este texto está dictada por consideraciones pedagógicas y no refleja necesariamente el orden de evolución. Pero la historia de la comunicación eléctrica es interesante e informativa, aunque sólo nos fijáramos en los significativos hechos pasados, cuando éstos ocurrieron y los nombres a ellos asociados.

Como es imposible hacer aquí una historia completa del tema, pues ello requeriría un libro particular,† presentaremos la siguiente relación cronológica de los más importantes inventos, descubrimientos y documentos. La cronología pro-

† Por ejemplo, Still (1946) o Black (1953 , Cap. 1); véase también *Scientific American*, Vol. 227, septiembre de 1972, el cual está dedicado completamente a la comunicación en muchos aspectos

porciona escasas indicaciones sobre la importancia e interrelación de estos hechos, pero por medio de discusiones adicionales de los últimos capítulos, será posible captar su completo significado. El lector puede, por lo tanto, percatarse de la utilidad de esta sección y consultarla de vez en cuando.

1800-1837	**Descubrimientos preliminares** Volta descubre la batería primaria; los tratados matemáticos de Fourier, Cauchy y Laplace; experimentos con electricidad y magnetismo por Oersted, Ampere, Faraday y Henry; la ley de Ohm (1826); primeros sistemas telegráficos por Gauss y Weber y por Wheatstone y Cooke.
1838-1866	**El nacimiento de la telegrafía** Morse perfecciona su sistema con la ayuda de Gale, Henry y Vail; Steinheil encuentra que la tierra puede ser empleada como un conductor; se inició el servicio comercial (1844); se inventó la técnica de la multicanalización; William Thomson (Lord Kelvin) calcula la respuesta a los pulsos de una línea telegráfica (1855); instalaron los cables transatlánticos por Cyrus Field y socios.
1845	Son enunciadas las leyes de Kirchhoff para circuitos.
1864	"Una Teoría Dinámica del Campo Electromagnético", por James Clerk Maxwell, predice la radiación electromagnética.
1876-1899	**El nacimiento de la telefonía** Es perfeccionado el transductor acústico por Alexander Graham Bell, después de varios intentos hechos por Reis; primer intercambio telefónico, en New Haven, Conn., con ocho líneas (1878); transductor de carbón, a botón de Edison; se introdujeron los circuitos de cables; Strowger inventa la conmutación paso a paso (1887); la teoría del cable cargado por Heaviside, Pupin y Campbell.
1887-1907	**La telegrafía inalámbrica** Heinrich Hertz comprueba la teoría de Maxwell; demostraciones por Marconi y Popov; Marconi patenta un sistema completo de telegrafía inalámbrica (1897); Sir Oliver Lodge desarrolla la teoría de los circuitos sintonizados; principia el servicio comercial que incluye sistemas barco-tierra y transatlánticos.
1892-1899	Publicaciones de Oliver Heaviside sobre cálculo operacional, circuitos y teoría electromagnética.
1904-1920	**La electrónica aplicada al radio y al teléfono** Lee De Forest inventa el Audión (triodo) basado en el diodo de Fleming; tipos básicos de filtros desarrollados por G. A. Campbell y otros; experimentos de radiotransmisión de AM; línea telefónica transcontinental con repetidores electrónicos con el Sistema Bell (1915); nace la telefonía multicanal sobre portadora; E. H. Armstrong perfecciona la radio receptor superheterodino (1918); primera estación radiodifusora, KDKA, en Pittsburgh.
1920-1928	Documentos guía sobre la teoría de la transmisión de señales y del ruido por J. R. Carson, H. Nyquist, J. B. Johnson y R. V. L. Hartley.
1923-1938	**El nacimiento de la televisión** Sistemas mecánicos de formación de imagen demostrados por Baird y Jenkins; análisis teórico de los requisitos de ancho de banda por Gray, Horton y Mathes; Farnsworth y Zworykin proponen sistemas electrónicos; tubos al vacío de rayos catódicos perfeccionados por DuMont y otros; empiezan las pruebas de campo y la transmisión experimental.
1931	Se inicia el servicio de teleimpresores.
1934	H. S. Black desarrolla el amplificador con retroalimentación negativa.
1936	"Un método para la reducción de perturbaciones en la emisión de señales de radio por un Sistema de Modulación en Frecuencia" por Armstrong, propicia la creación de la radio de FM.
1937	Alec Reeves concibe la modulación por codificación de pulsos.
1938-1945	**Segunda Guerra Mundial** Se desarrollan los sistemas de radar y microondas; la FM es usada en forma amplia en las comunicaciones militares; mejoras en electrónica, hardware y teoría en todas las áreas; Weiner y Kolmogoroff aplican métodos estadísticos a los problemas de detección de señales.

1948	C. E. Shannon publica "Una Teoría Matemática de la Comunicación".
1948-1951	El transistor es inventado por Bardeen, Brattain y Shockley.
1950	El sistema multicanal por división en tiempo es aplicado a la telefonía.
1955	J. R. Pierce propone sistemas para comunicación por satélites.
1956	Primer cable telefónico transoceánico (36 canales de voz).
1958	Se desarrollan sistemas de transmisión de datos a larga distancia con propósitos militares.
1960	Maiman demuestra el primer laser.
1961	Los circuitos integrados entran en la producción comercial.
1962	Principia la comunicación por satélite con el Telstar I.
1962-1966	**El nacimiento de la comunicación digital de alta velocidad** El servicio de transmisión de datos aparece en forma comercial; canales de banda ancha para señalización digital; la modulación por codificación de pulsos se hace factible en la transmisión de voz y TV; mejoras en la teoría e implementación de la transmisión digital; incluyendo métodos de codificación para control de error por Bose, Chaudhuri, Wozencraft y otros, y el desarrollo de la igualación para adaptación por Lucky y colaboradores.
1963	Osciladores de microondas de estado sólido perfeccionados por Gunn y otros.
1964	Se ponen en servicio los sistemas de conmutación telefónica completamente electrónicos (No. 1 ESS).
1965	El Mariner IV transmite imágenes de Marte a la Tierra.
1966-1975	**Precursores del futuro** Sistemas de TV por cable; enlaces experimentales de comunicación por laser; ensayos en el campo de la fonovisión; desarrollo de trabajos en las fibras ópticas, guías de onda helicoidales; filtros digitales, dispositivos acoplados en carga y circuitos integrados en gran escala.

1.6 PLAN DEL LIBRO

Este estudio de sistemas de comunicación comienza con las limitaciones físicas fundamentales, su descripción y análisis. En el capítulo 2 se repasan las herramientas matemáticas de la teoría de la señal y sistemas lineales, recalcando el acceso al dominio de la frecuencia que conduce a los conceptos de espectro y ancho de banda. De todas maneras, éste será un campo familiar, basado en métodos de transformación elementales y de la teoría de los circuitos. El capítulo 3 es un curso muy corto sobre probabilidad y estadísticas aplicadas a las señales particulares aleatorias de interés en la comunicación. Nuestra tarea principal será describir el ruido indeseable, particularmente desde el punto de vista espectral.

Teniendo desplegada la herramienta necesaria, enfoquemos la materia de ingeniería de la comunicación, o sea, la teoría y práctica de la transmisión de la señal. El capítulo 4 considera los parámetros básicos del sistema y estrategias de diseño en la transmisión de banda base, es decir, sin modulación. La modulación de onda continua se trata en los capítulos 5 al 7, mientras que en el capítulo 8 se explican la teoría del muestreo y la modulación de pulsos. En estos capítulos se considerará el problema de la recuperación de la señal en presencia de ruido e interferencia. Las técnicas de la multicanalización, la reducción del ruido de banda ancha y efectos de umbral, se examinan cuando tienen aplicación. En la mayor parte, los elementos del sistema e instrumentación serán descritos como "cajas negras" con propiedades específicas, pero en ocasiones se quitará su tapa para ver qué contienen y cómo se efectúan las diversas funciones en la práctica.

La teoría de la información y sus implicaciones en la comunicación eléctrica están bosquejadas en el capítulo 9. Es de particular importancia la ecuación de Hartley-Shannon, la cual expresa en forma cuantitativa las limitaciones de ancho de banda y ruido en la transmisión de la información. A la luz de esta teoría, es posible obtener algunas conclusiones acerca de los méritos relativos de los sistemas convencionales y obtener algunas indicaciones para poder diseñar mejores sistemas.

El texto concluye, en el capítulo 10, con la transmisión digital de datos, el área de más rápida expansión en la ingeniería de la comunicación. Este tema no sólo incluye todo el material previo, sino que también da un significado adicional a la teoría matemática de la comunicación.

Cada capítulo contiene varios ejercicios elaborados para aclarar y reforzar los conceptos y las técnicas analíticas conforme han sido consideradas. Así, los estudiantes podrán probar la comprensión del material al trabajar sobre estos ejercicios. Las respuestas se proporcionan donde se considera necesario.

Ciertos temas opcionales o más avanzados se han diseminado en el texto. Estas secciones están identificadas por el símbolo ★ y pueden ser omitidas sin que haya serias fallas de continuidad. Al final de los tres Apendices del libro, ha sido insertado otro material de naturaleza suplementaria.

También al final del libro el lector encontrará varias tablas y una lista de lecturas suplementarias seleccionadas. Las primeras contienen muchas de las relaciones matemáticas y datos numéricos para resolver los problemas del final de los capítulos. Las segundas son una bibliografía con anotaciones, de libros y documentos en beneficio de quienes deseen analizar un tema con mayor profundidad.

2

SEÑALES, ESPECTROS Y FILTROS

Las señales en la comunicación eléctrica son magnitudes variables en el tiempo, tales como el voltaje y la corriente. La descripción de una señal $v(t)$ usualmente existe en el *dominio del tiempo*, donde la variable independiente es t. Pero para el trabajo de comunicaciones, a menudo es más conveniente describir las señales en el *dominio de la frecuencia*, donde la variable independiente es f. En forma burda, podemos decir que la función del tiempo está compuesta de un cierto número de componentes de frecuencia, cada una con amplitud y fase apropiadas. Así, si la señal existe físicamente en el dominio del tiempo, podemos decir que consiste de aquellas componentes en su descripción del dominio de la frecuencia, llamado el *espectro*.

El análisis espectral, basado en las series y transformadas de Fourier, es una herramienta poderosa en la ingeniería de la comunicación. En consecuencia, centraremos nuestra atención en la teoría de Fourier antes que en otras técnicas como las transformadas de Laplace y el análisis del dominio del tiempo. Existen varias razones para recalcar esto:

Primero, el dominio de la frecuencia es esencialmente un punto de vista de estado permanente y, para muchos propósitos, es razonable restringir nuestra atención al comportamiento de estado permanente de un sistema de comunicación. Realmente, teniendo en cuenta la multitud de señales posibles que un sistema puede manejar, sería una tarea imposible encontrar soluciones detalladas de los transitorios de cada una de ellas. Segundo, el acceso espectral nos permite considerar clases

completas de señales que posean propiedades similares en el dominio de la frecuencia. Esto no sólo nos lleva a un conocimiento más profundo del análisis, sino que es de gran valor para el diseño. Es muy improbable, por ejemplo, que una técnica tan importante como la modulación de banda lateral única haya podido desarrollar sin la ayuda de los conceptos espectrales. Tercero, muchos componentes de un sistema de comunicación se pueden clasificar como dispositivos lineales e invariantes en el tiempo; siendo así, podemos describirlos por sus *características de respuesta en frecuencia* las cuales, a su vez, facilitan aún más el análisis y el trabajo de diseño.

Por lo tanto, este capítulo está dedicado a la revisión† y elaboración del análisis de Fourier en señales y características de respuesta de frecuencia de componentes del sistema, principalmente aquellas componentes selectivas en frecuencia conocidas como filtros. Sin embargo, no debemos pensar que el enfoque espectral sea el único método empleado por los ingenieros de la comunicación. Existen algunos problemas donde no se puede aplicar directamente, y otros donde es más conveniente emplear diferente técnica. Por lo mismo, cada nuevo problema debe ser atacado con mente clara y un buen conjunto de herramientas analíticas.

Como primer paso, en la mayor parte de nuestro trabajo, escribiremos ecuaciones que representen señales o componentes. Pero no olvidamos, que tales ecuaciones son solamente *modelos* matemáticos de entidades físicas, por lo general modelos imperfectos. En efecto, una descripción completamente fiel de las señales o componentes más simples sería prohibitiva por compleja en forma matemática y, en consecuencia, nada útil para propósitos de ingeniería. Por lo tanto, los modelos que se precisan serán los que representen, con la mínima complejidad, aquellas propiedades pertinentes al problema que se confronta. Esto, algunas veces nos conducirá a construir varios modelos diferentes de acuerdo a las necesidades. Así, dado un problema particular, la elección del modelo que debamos usar estará basada en la comprensión de los fenómenos físicos involucrados y las limitaciones de las matemáticas; en pocas palabras, es ingeniería.

2.1 SEÑALES DE C.A. Y REDES

La figura 2.1 representa una red eléctrica de dos puertos; un sistema impulsado por una señal de entrada $x(t)$, la cual da lugar a una señal de salida $y(t)$. Se dice que $x(t)$ es una señal senoidal o de C.A. si

$$x(t) = A_x \cos(\omega_0 t + \theta_x) \qquad -\infty < t < \infty \tag{1}$$

donde A_x es la *amplitud* (en volts o amperes), ω_0 es la *frecuencia angular* (en radianes por segundo) y θ_x es la *fase* (en radianes o grados). Es también conveniente introducir la *frecuencia cíclica* $f_0 = \omega_0/2\pi$ (en ciclos por segundo o hertz) así como el *pe-*

† Se supone que el lector tiene bases acerca de la teoría de la transformada y de los sistemas lineales. En consecuencia, se han omitido algunas demostraciones y deducciones.

Entrada
$x(t) = A_x \cos(\omega_0 t + \theta_x)$
Sistema
Salida
$y(t) = A_y \cos(\omega_0 t + \theta_y)$

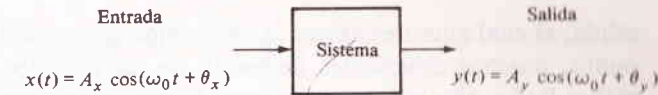



FIGURA 2.1
Sistema en la condición de CA de estado permanente.

riodo (en segundos), el cual es $T_0 = 1/f_0 = 2\pi/\omega_0$. El significado de T_0 es que $x(t)$ se repite cada T_0 segundo todo el tiempo. Obviamente, la señal permanece real siempre, pero la ecuación (1) es una representación o modelo útil y conveniente para una señal senoidal de duración finita si la duración es mucho mayor que el periodo.

Ahora, supongamos que la red es *lineal*, *invariante en el tiempo* y estable asintóticamente. En esencia, esto significa que acepta la superposición, no hay parámetros variables con el tiempo y el comportamiento natural decae con éste. Bajo estas condiciones, la señal de salida será también senoidal y de la misma frecuencia que la de entrada, difiriendo sólo en amplitud y fase, es decir,

$$y(t) = A_y \cos(\omega_0 t + \theta_y) \qquad -\infty < t < \infty \tag{2}$$

Por lo tanto, dados los parámetros de la señal de entrada y las características de la red, sólo necesitamos resolver para A_y y θ_y para describir completamente la señal de salida resultante. Éste es, por supuesto, el conocido problema de la C.A. en estado permanente.

Es bien conocido que tales problemas se resuelven más fácilmente haciendo uso de las funciones exponenciales del tiempo, de la forma $e^{j\omega t}$, más que de las funciones senoidales. Veamos el caso, por ejemplo, de un circuito con impedancia compleja $Z(j\omega)$; si la corriente a través del circuito es $e^{j\omega t}$ entonces, por la definición de impedancia, el voltaje resultante es simplemente $Z(j\omega)e^{j\omega t}$. De manera similar, para el caso de la red de la figura 2.1, definimos la *función de transferencia* de la red $H(j\omega)$ tal que $y(t) = H(j\omega)e^{j\omega t}$ cuando $x(t) = e^{j\omega t}$; esto es,

$$H(j\omega) \triangleq \frac{y(t)}{x(t)} \qquad \text{cuando } x(t) = e^{j\omega t} \tag{3}$$

una definición† que generalizaremos en lo sucesivo. Combinando la ecuación (3) con el principio de la superposición, se sigue que si $x(t)$ es una combinación lineal de exponenciales, digamos,

$$x(t) = \alpha_1 e^{j\omega_1 t} + \alpha_2 e^{j\omega_2 t} + \cdots \tag{4a}$$

entonces

$$y(t) = H(j\omega_1)\alpha_1 e^{j\omega_1 t} + H(j\omega_2)\alpha_2 e^{j\omega_2 t} + \cdots \tag{4b}$$

donde α_1 y α_2 son constantes, $H(j\omega_1)$ representa $H(j\omega)$ evaluada en $\omega = \omega_1$, etc.

Sin embargo, aún no hemos resuelto el problema de la C.A. permanente en la

† El símbolo $\triangleq$ significa "es igual por definición".

salida, el cual ahora establece la conversión de senoidales a exponenciales. En este punto, muchos estudiantes de teoría de los circuitos, probablemente escribirían $x(t) = \text{Re}\,[A_x e^{j\theta_x} e^{j\omega_0 t}]$, pero preparando el camino para desarrollos futuros, preferimos una expresión como la de la ecuación (4*a*). Con este fin recurrimos a un corolario del teorema de Euler,† es decir:

$$\cos\phi = \tfrac{1}{2}(e^{j\phi} + e^{-j\phi}) \qquad (5)$$

Así, la señal senoidal $x(t)$ de la ecuación (1) se puede escribir ahora

$$x(t) = A_x \tfrac{1}{2}[e^{j(\omega_0 t + \theta_x)} + e^{-j(\omega_0 t + \theta_x)}]$$
$$= \frac{A_x}{2} e^{j\theta_x} e^{j\omega_0 t} + \frac{A_x}{2} e^{-j\theta_x} e^{-j\omega_0 t} \qquad (6a)$$

y puesto que esta tiene la misma forma de la ecuación (4*a*), la sustitución apropiada en la ecuación (4*b*) da

$$y(t) = H(j\omega_0)\frac{A_x}{2} e^{j\theta_x} e^{j\omega_0 t} + H(-j\omega_0)\frac{A_x}{2} e^{-j\theta_x} e^{-j\omega_0 t} \qquad (6b)$$

Como éste es un resultado correcto, se puede ordenar y hacerlo más comprensible por medio de los siguientes dos pasos.

Primero, como la función de transferencia es en general una cantidad compleja, la expresaremos en la forma polar

$$H(j\omega) = |H(j\omega)|\, e^{j\,\arg[H(j\omega)]} \qquad (7)$$

donde $H(j\omega)$ es la *magnitud* y $[H(j\omega)]$ es el ángulo.‡ Segundo, haciendo caso omiso del hecho de que $H(j\omega)$ es compleja, $y(t)$ debe ser una función real del tiempo, puesto que $x(t)$ es real; y esto será si y sólo si

$$H(-j\omega) = H^*(j\omega)$$
$$= |H(j\omega)|\, e^{-j\,\arg[H(j\omega)]} \qquad (8)$$

donde $H^*(j\omega)$ es el conjugado complejo de $H(j\omega)$. La relación conjugada compleja tiene validez, de hecho, para cualquier red real.

† El teorema de Euler es la más versátil de las identidades trigonométricas. En la tabla B, al final del libro, se da una breve lista de estas útiles relaciones.

‡ O sea, que si H_r y H_i son las partes real e imaginaria de $H(j\omega)$, respectivamente, entonces

$$|H(j\omega)| = \sqrt{H_r^2 + H_i^2} \qquad \arg[H(j\omega)] = \arctan\frac{H_i}{H_r}$$

Se aconseja al lector que tenga dificultades con estos desarrollos, repasar el tema de números complejos.

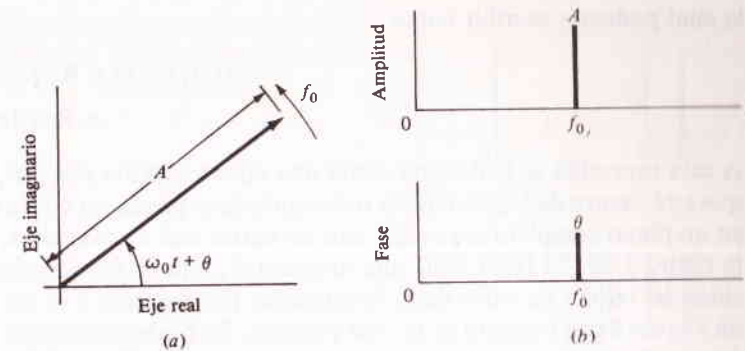



FIGURA 2.2
Representación de $A\cos(\omega_0 t + \theta)$. *a*) diagrama fasorial; *b*) espectros de líneas.

Aplicando las ecuaciones (7) y (8) a la (6*b*) y simplificando, se llega a

$$y(t) = |H(j\omega_0)|\frac{A_x}{2}[e^{j(\omega_0 t + \theta_x + \arg[H(j\omega_0)])} + e^{-j(\omega_0 t + \theta_x + \arg[H(j\omega_0)])}]$$
$$= \underbrace{|H(j\omega_0)|A_x}_{A_y}\cos(\omega_0 t + \underbrace{\theta_x + \arg[H(j\omega_0)]}_{\theta_y}) \qquad (9)$$

en la cual se identifica a

$$A_y = |H(j\omega_0)|A_x \qquad \theta_y = \theta_x + \arg[H(j\omega_0)] \qquad (10)$$

Así, tenemos estas ecuaciones muy simples para la respuesta de amplitud y fase de salida de la C.A. de estado permanente, con tal que conozcamos la función de transferencia de la red. Esto se tratará con más detalle después que hayamos considerado la interpretación del dominio de la frecuencia.

EJERCICIO 2.1 Escribir $H(j\omega_0)$ y $(-j\omega_0)$ en forma polar y demostrar que la parte imaginaria de la ecuación (6*b*) no es cero cuando $H(-j\omega_0) \neq H^*(j\omega_0)$.

Fasores y espectro de líneas

Además de facilitar el análisis de redes, la conversión de funciones senoidales a exponenciales también fundamenta la noción de dominio de la frecuencia por medio de diagramas fasoriales. Para establecer esta idea, consideremos la función sinusoide arbitraria

$$v(t) = A\cos(\omega_0 t + \theta) \qquad \omega_0 = 2\pi f_0$$

la cual podemos escribir como

$$A \cos(\omega_0 t + \theta) = \text{Re}\,[Ae^{j(\omega_0 t+\theta)}]$$
$$= \text{Re}\,[Ae^{j\theta}e^{j\omega_0 t}] \qquad (11)$$

A esta expresión se le designa como una *representación fasorial*, ya que el término que está dentro de los paréntesis rectangulares se puede ver como un vector giratorio en un plano complejo cuyos ejes son las partes real e imaginaria, como se ilustra en la figura 2.2*a*. El fasor tiene una longitud A, gira en sentido contrario a las manecillas del reloj a un valor de f_0 revoluciones por segundo y en un tiempo $t = 0$ hace un ángulo θ con respecto al eje real positivo. En cualquier tiempo t la proyección del fasor sobre el eje real, es decir, su parte real, es igual a la de la sinusoide $v(t)$.

Obsérvese que solamente se necesitan tres parámetros para especificar a un fasor: la amplitud, la fase relativa y la frecuencia de rotación. Al describir el mismo fasor en el *dominio de la frecuencia*, vemos que se define sólo para la frecuencia particular f_0. Debemos asociar con esta frecuencia las correspondientes amplitud y fase. Por lo tanto, una descripción adecuada del dominio de la frecuencia sería el *espectro de líneas* de la figura 2.2*b*, el cual consiste de dos gráficas, amplitud contra frecuencia y fase contra frecuencia. No obstante que la figura 2.2*b* parece simple al grado de ser trivial, posee un gran valor conceptual, especialmente cuando se aplica a señales más complicadas. Pero antes de adoptar este paso, se deben establecer cuatro normas convencionales que se usan en la construcción de espectros de líneas.

1. En todos nuestros dibujos espectrales, la variable independiente será la *frecuencia cíclica f* en hertz, en lugar de la frecuencia en radianes ω, o cualquier frecuencia específica tal como f_0, que sea constante, se identificará con un subíndice. Usaremos, sin embargo, ω con o sin subíndice como una notación corta para $2\pi f$ en varias ecuaciones, puesto que ocurre esa combinación muy a menudo.
2. Los ángulos de fase se medirán con respecto a ondas *cosenoidales* o equivalentes, con relación al eje real positivo del diagrama fasorial. Por lo que las ondas senoidales deben convertirse a cosenoidales por medio de la identidad

$$\text{sen}\,\omega t = \cos(\omega t - 90°) \qquad (12)$$

3. Consideraremos a la amplitud como una cantidad siempre *positiva*; cuando aparece un signo negativo, o varios, deben introducirse en la fase, por ejemplo:

$$-A \cos \omega t = A \cos(\omega t \pm 180°) \qquad (13)$$

y no importa si se toma $+180°$ o $-180°$, puesto que el fasor finaliza en el mismo lugar por uno u otro camino.

4. Los ángulos de fase por lo general se expresan en grados, aunque otros ángulos estén inherentemente en radianes, por ejemplo, ωt en las ecuaciones (12) y (13) están expresadas en radianes, mientras que $-90°$ y $\pm$ 180° obviamente están en grados. No debe resultar ninguna confusión de esta notación combinada, puesto que los ángulos expresados en grados siempre llevarán el símbolo apropiado.

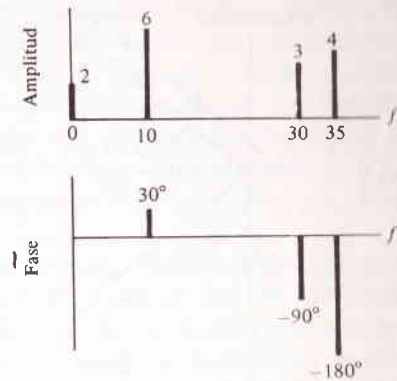



FIGURA 2.3
El espectro de líneas de frecuencia positiva.

Para ilustrar estas convenciones y llevar la idea del espectro de líneas en una forma más amplia, supongamos una señal formada de una suma de sinusoides, como

$$w(t) = 2 + 6\cos(2\pi 10t + 30°) + 3\,\text{sen}\,2\pi 30t - 4\cos 2\pi 35t$$

Convirtiendo el término constante (C.D.) como una sinusoide de frecuencia cero y aplicando las ecuaciones (12) y (13), se tiene

$$w(t) = 2\cos 2\pi 0t + 6\cos(2\pi 10t + 30°) + 3\cos(2\pi 30t - 90°) + 4\cos(2\pi 35t - 180°)$$

así, el espectro de la figura 2.3 tiene sus líneas de amplitud y fase en 0, 10, 30 y 35 Hz.

Las figuras 2.2*b* y 2.3 conocidas como *espectros de líneas de un lado* o *de frecuencia positiva*, se pueden generar con cualquier combinación lineal de sinusoides. Pero hay otra representación espectral, la cual es sólo ligeramente más complicada y viene a ser mucho más útil. Se forma escribiendo una sinusoide como la suma de dos exponenciales, similar a la ecuación (6*a*), es decir,

$$A\cos(\omega_0 t + \theta) = \frac{A}{2}e^{j\theta}e^{j\omega_0 t} + \frac{A}{2}e^{-j\theta}e^{-j\omega_0 t} \qquad (14)$$

la cual designaremos como la representación *fasorial conjugada*, puesto que los dos términos son complejos conjugados uno del otro. El diagrama correspondiente se muestra en la figura 2.4*a*, donde se tienen ahora dos fasores con iguales longitudes, pero con ángulos y direcciones de rotación opuestos. Por lo tanto, la figura ilustra la suma vectorial en cualquier tiempo, la cual es igual a $v(t)$. Nótese que la suma siempre cae sobre el eje real —cual debe ser, puesto que $v(t)$ es real.

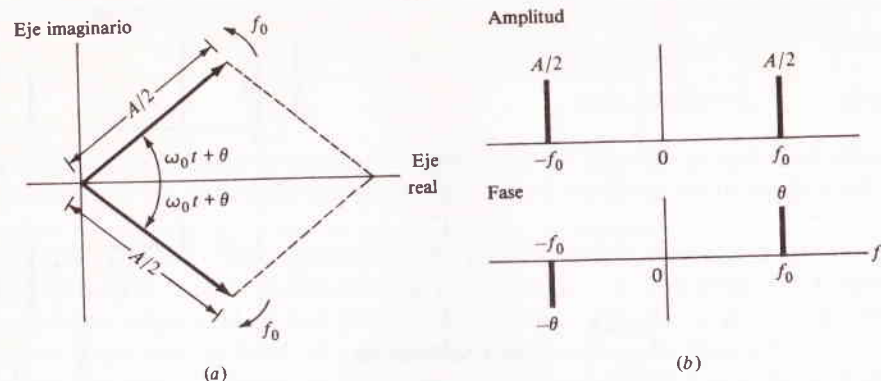



FIGURA 2.4
a) Representación fasorial conjugada de A cos $(\omega_0 t + \theta)$; b) espectro de líneas de doble lado.

Un espectro de líneas tomado de la expresión fasorial conjugada, debe incluir las frecuencias negativas para indicar las dos direcciones de rotación. Así, en el *espectro de líneas de dos lados* de la figura 2.4*b*, la mitad de la amplitud original queda asociada con cada una de las dos frecuencias $\pm f_0$. Las normas de construcción de tales espectros son muy simples; una poca de observación nos mostrará que las líneas de amplitud tienen simetría par en f, mientras que las líneas de fase tienen simetría impar. La simetría es una consecuencia directa de que $v(t)$ es una función del tiempo real (no compleja). Así, por ejemplo, la figura 2.5 es la versión de dos lados de la figura 2.3.

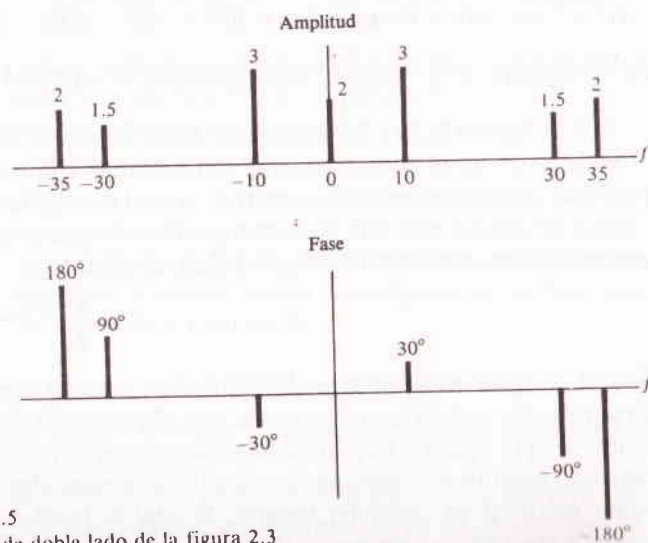



FIGURA 2.5
La versión de doble lado de la figura 2.3

Existe una diferencia significativa en la interpretación del espectro de líneas de doble lado comparada con el espectro de frecuencia positiva. En éste, una línea sencilla representa una onda cosenoidal dada por Re$[e^{j\omega t}]$. Pero en el caso del doble lado, una línea representa por sí misma un fasor sencillo, y para obtener una función real del tiempo se requiere el término conjugado. Así, siempre que hablemos de algún intervalo de frecuencia en un espectro de doble lado, tal como el de f_1 a f_2, debemos incluir el correspondiente intervalo de frecuencia negativa, $-f_1$. f_2—. Una Una notación simple especificando ambos intervalos es $f_1 \leq |f| \leq f_2$.

Debe recalcarse que estos espectros de líneas son sólo formas gráficas de representar ciertas señales en términos de la forma $e^{j\omega t}$. El efecto de una red sobre una señal, queda reducido entonces al problema de la C.A. de estado permanente. También debe puntualizarse que en algunos casos, el espectro de amplitud es más importante que el espectro de fase. Ambas partes son necesarias, por supuesto, para definir sin ambigüedad a una señal en el dominio del tiempo, pero el espectro de amplitud por sí mismo nos muestra cuáles frecuencias están y en qué proporciones, o sea, nos dice el *contenido de frecuencia* de la señal. La ventaja específica de la versión de doble lado, se pondrá de manifiesto conforme avancemos.

EJERCICIO 2.2 Explique por qué en la figura 2.5 el término de frecuencia cero está representado por una sola línea de amplitud y diga cómo cambiaría si el término fuera negativo.

Funciones de transferencia y respuesta en frecuencia

Así como una señal se puede describir en el dominio de la frecuencia por medio de su espectro, una red o sistema se puede describir en términos de sus características de respuesta en frecuencia determinadas a partir de la función de transferencia $H(j\omega)$. Al recalcar el punto de vista del dominio de la frecuencia, es conveniente introducir ahora una nueva notación $H(f)$, definida por

$$H(f) \triangleq H(j\omega) \qquad \text{con } \omega = 2\pi f \tag{15}$$

la cual, en lo sucesivo se designará como la función de respuesta en frecuencia o como la función de transferencia. Aunque la ecuación (15) viola ligeramente la notación formal matemática, ello significa simplemente que $H(f)$ es idéntica a $H(j\omega)$ con ω sustituida por $2\pi f$. De la ecuación (8) se tiene, para una red real,

$$H(-f) = H^*(f) \tag{16a}$$

o en forma polar,

$$|H(-f)| = |H(f)| \qquad \arg[H(-f)] = -\arg[H(f)] \tag{16b}$$

una propiedad conocida como *simetría hermitiana*.

Para la interpretación de $H(f)$, supongamos que la señal de entrada es un fasor sencillo

$$x(t) = A_x e^{j\theta_x} e^{j\omega_0 t} \qquad \omega_0 = 2\pi f_0$$

así, de la ecuación (4*b*), la salida es

$$y(t) = H(f_0) A_x e^{j\theta_x} e^{j\omega_0 t}$$

$$= \underbrace{|H(f_0)| A_x}_{A_y} \exp j\underbrace{(\theta_x + \arg [H(f_0)])}_{\theta_y} e^{j\omega_0 t}$$

Por lo tanto, igual que en el caso de la C.A., los parámetros de las señales de entrada y salida están relacionados por

$$\frac{A_y}{A_x} = |H(f_0)| \qquad \theta_y - \theta_x = \arg [H(f_0)] \tag{17}$$

donde $|H(f_0)|$ es la relación de amplitudes y $[H(f_0)]$ es la diferencia de fase, ambas a la frecuencia específica f_0. Generalizando, concluimos que $|H(f)|$ da la *relación de amplitudes* del sistema (designada algunas veces como la *respuesta de amplitud* o *ganancia*) y $[H(f)]$ da el corrimiento de fase, ambos como funciones continuas de la frecuencia. Las gráficas de estas dos relaciones contra f dan una representación en el dominio de la frecuencia del sistema, análoga al espectro de amplitud y fase de una señal. Más aún, la simetría hermitiana de la ecuación (16) significa que la relación de amplitudes será una función par de la frecuencia mientras que el cambio o corrimiento de fase será una función impar. En seguida presentamos un ejemplo ilustrativo, después de considerar brevemente el problema de la determinación de $H(f)$ para un sistema particular.

Dado el diagrama del circuito de una red, encontrar $H(f)$ no es otra cosa que un problema de análisis fasorial, es decir, suponemos que la entrada es $e^{j2\pi ft}$ y calculamos la salida, la cual será $H(f)e^{j2\pi ft}$. Podemos recurrir a todas las herramientas de la ingeniería eléctrica (Ley de Ohm para impedancia compleja, leyes de corrientes de Kirchoff, etc.). En forma alternativa, si un sistema se describe por medio de una ecuación diferencial lineal con coeficientes constantes, de la forma general

$$a_n \frac{d^n y}{dt^n} + \cdots + a_1 \frac{dy}{dt} + a_0 y(t) = b_m \frac{d^m x}{dt^m} + \cdots + b_1 \frac{dx}{dt} + b_0 x(t) \tag{18a}$$

entonces

$$H(f) = \frac{b_m (j2\pi f)^m + \cdots + b_1 (j2\pi f) + b_0}{a_n (j2\pi f)^n + \cdots + a_1 (j2\pi f) + a_0} \tag{18b}$$

cuya derivación se deja como ejercicio. Un tercer método que implica la respuesta al impulso del sistema, se presenta en la Secc. 2.5.

EJERCICIO 2.3 Deduzca la ecuación (18*b*) sustituyendo $x(t) = e^{j2\pi ft}$ y $y(t) = H(f)e^{j2\pi ft}$ en la ecuación (18*a*) y resuelva para $H(f)$.

Ejemplo 2.1 Filtro RC pasabajas

La red de la figura 2.6*a*, que es virtualmente clásica en comunicaciones, se conoce como un filtro RC pasabajas; $x(t)$ es el voltaje de entrada y $y(t)$ es el voltaje de salida bajo condiciones de circuito abierto (sin carga). Encontrar la función de transferencia es algo sencillo, puesto que con $x(t) = e^{j2\pi ft}$, la aplicación de la relación del divisor de voltaje, lleva a

$$y(t) = \frac{Z_C}{R + Z_C} e^{j2\pi ft}$$

donde la impedancia del capacitor es $Z_c = 1/j2\pi fC$. Así,

$$H(f) = \frac{(1/j2\pi fC)}{R + (1/j2\pi fC)} = \frac{1}{1 + j2\pi RCf}$$

$$= \frac{1}{1 + j(f/B)} \tag{19}$$

en la cual hemos definido el parámetro del sistema

$$B \triangleq \frac{1}{2\pi RC}$$

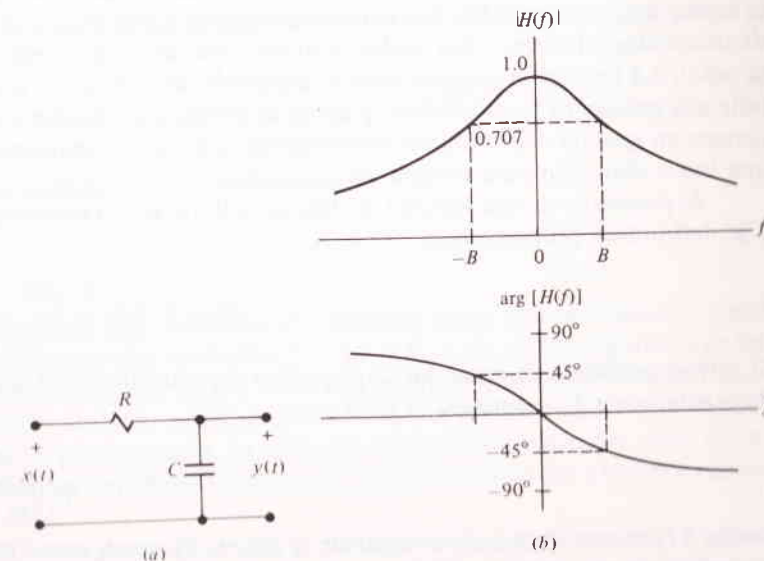



FIGURA 2.6
Filtro pasabajas RC. *a*) Circuito; *b*) función de transferencia.

La conversión a forma polar da la relación de amplitudes y el corrimiento de fase como

$$|H(f)| = \frac{1}{\sqrt{1 + (f/B)^2}} \tag{20a}$$

$$\arg [H(f)] = -\arctan \frac{f}{B} \tag{20b}$$

lo cual está expresado en la gráfica en la figura 2.6*b*. ////

EJERCICIO 2.4 Supóngase que se intercambian la resistencia y el capacitor en la figura 2.6*a*. Encontrar la nueva $H(f)$ y bosquejando $|H(f)|$, justificar que a este circuito se le llame filtro RC *pasaaltas*.

2.2 SEÑALES PERIÓDICAS Y SERIES DE FOURIER

Se dice que una señal $v(t)$ es periódica con un periodo de repetición T_0, si, para cualquier entero m,

$$v(t \pm mT_0) = v(t) \qquad -\infty < t < \infty \tag{1}$$

Puesto que esto implica una señal que continúa en forma indefinida, resulta conveniente tener siempre presente la diferencia entre una señal y su modelo matemático. Suponiendo que la ecuación (1) es un modelo razonable, entonces se puede recurrir a la expansión en serie de Fourier para descomponer a $v(t)$ en una combinación lineal de sinusoides o fasores —los cuales, a su vez, nos conducen al espectro de líneas de la señal. La condición esencial para la expansión de Fourier es que $v(t)$ tenga una potencia promedio bien definida, y como la potencia promedio y los promedios de tiempo en general son términos comúnmente usados en comunicaciones, haremos una breve digresión para normar los conceptos.

El *promedio* de una función del tiempo arbitraria $v(t)$ se denotará por $<v(t)>$ y se definirá en general como

$$\langle v(t)\rangle \triangleq \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} v(t)\, dt \tag{2}$$

Si $v(t)$ es periódica, la ecuación (2) se reduce al promedio sobre cualquier intervalo de exactamente T_0 segundos, es decir,

$$\langle v(t)\rangle = \frac{1}{T_0}\int_{T_0} v(t)\, dt \tag{3}$$

donde $\int T_0$ se establece para integrar de t_1 a $t_1 + T_0$ con t_1 como una constante arbitraria. La potencia promedio es, por supuesto, el promedio de tiempo de la potencia instantánea, por ejemplo, si $v(t)$ es el voltaje a través de una resistencia de 1 ohm

(Ω), la potencia instantánea es $v^2(t)$ y la potencia promedio es $<v^2(t)>$. Para nuestros propósitos es conveniente suponer que todas las resistencias están normalizadas a la unidad, de tal manera que si $v(t)$ es un voltaje o una corriente periódica, su *potencia promedio P* se define como

$$P \triangleq \langle |v(t)|^2\rangle = \frac{1}{T_0}\int_{T_0} |v(t)|^2\, dt \tag{4}$$

Se dice entonces que la señal $v(t)$ tiene una potencia promedio bien definida, si existe la integral (4) y da lugar a una cantidad finita. Nótese también que se ha permitido la posibilidad de señales complejas escribiendo $|v(t)|^2 = v(t)v^*(t)$ en lugar de $v^2(t)$.

EJERCICIO 2.5 Emplee las ecuaciones (3) y (4) para demostrar que si

$$v(t) = A\cos(\omega_0 t + \theta) \qquad \omega_0 = \frac{2\pi}{T_0} \tag{5a}$$

entonces

$$\langle v(t)\rangle = 0 \qquad \langle |v(t)|^2\rangle = \frac{A^2}{2} \tag{5b}$$

Así, la potencia promedio de una sinusoide depende solamente de su amplitud.

Series de Fourier y espectros de líneas

La expansión de la *serie de Fourier exponencial* de una señal periódica $v(t)$ es

$$v(t) = \sum_{n=-\infty}^{\infty} c(nf_0)e^{j2\pi nf_0 t} \qquad f_0 = \frac{1}{T_0} \tag{6}$$

donde $c(nf_0)$ es el coeficiente de Fourier n-ésimo

$$c(nf_0) \triangleq \frac{1}{T_0}\int_{T_0} v(t)e^{-j2\pi nf_0 t}\, dt \tag{7}$$

La ecuación (6) establece que $v(t)$ puede ser expresada como una combinación lineal o suma gravada de fasores a las recuencias $f = nf_0 = 0, \pm f_0, \pm 2f_0$, el gravado de los factores lo da la ecuación (7). Si $v(t)$ tiene una potencia promedio bien definida, entonces la suma del segundo miembro de la ecuación (6) converge a $v(t)$ dondequiera que ella es finita y continua, lo cual sería verdad para cualquier señal física. En cuanto a propósitos de ingeniería concierne, podemos visualizar a las series como algo idéntico a $v(t)$.

En realidad, rara vez se desarrolla la suma indicada por la ecuación (6) para encontrar $v(t)$; en su lugar, dada una señal periódica, usamos la ecuación (7) para encontrar sus coeficientes de Fourier, y de eso, el espectro de líneas. Antes de

abordar ese aspecto, observemos un poco a $c(nf_0)$. En general, los coeficientes son cantidades complejas, siempre y cuando la señal sea real; por lo tanto, podemos escribir

$$c(nf_0) = |c(nf_0)| e^{j \arg[c(nf_0)]}$$

Si $v(t)$ es una función *real*, entonces reemplazando n por $-n$ en la ecuación (7) nos muestra que

$$c(-nf_0) = c^*(nf_0) \tag{8}$$

así, de nuevo tenemos fasores conjugados y simetría hermitiana. Finalmente nótese que $c(nf_0)$ no depende del tiempo, puesto que la ecuación (7) es una integral definida donde t es la variable de integración. Es útil, sin embargo, considerar a $c(nf_0)$ como una función de la frecuecia f definida solamente para las frecuencias discretas $f = nf_0$.

Volviendo a la interpretación espectral, vemos de la ecuación (6) que una señal periódica contiene sólo aquellos componentes de frecuencia que son *múltiplos enteros de la frecuencia fundamental* $f_0 = 1/T_0$ o, en otras palabras, todas las frecuencias son *armónicas* de la fundamental. Puesto que el coeficiente de la n-ésima armónica es $c(nf_0)$, su amplitud y fase son $|c(nf_0)|$ y $[c(nf_0)]$, respectivamente. Por lo tanto, tenemos un espectro de línea de dos lados con $|c(nf_0)|$ dando la amplitud y $[c(nf_0)]$ la fase. A continuación se listan algunas de las más importantes propiedades de tales espectros.

1 Todas las líneas espectrales están igualmente espaciadas por f_0, puesto que todas las frecuencias son armónicas en relación con la fundamental.
2 La componente de C.D. es igual al *valor promedio* de la señal, puesto que haciendo a $n = 0$ en la ecuación (7) se tiene

$$c(0) = \frac{1}{T_0} \int_{T_0} v(t)\, dt = \langle v(t) \rangle \tag{9}$$

Por lo que los valores calculados de $c(0)$ pueden ser corroborados observando a $v(t)$ —lo cual constituye una buena práctica, ya que la integración frecuentemente nos lleva a una forma indeterminada de $c(0)$.
3 Si $v(t)$ es *real*, el espectro de amplitud tiene simetría par, mientras que el espectro de fase tiene simetría impar, es decir,

$$|c(-nf_0)| = |c(nf_0)| \qquad \arg[c(-nf_0)] = -\arg[c(nf_0)] \tag{10}$$

lo cual procede de la ecuación (8).
4 Si una señal real tiene *simetría par* en el tiempo, tal que

$$v(-t) = v(t) \tag{11a}$$

entonces $c(nf_0)$ es completamente ***real*** y

$$\arg[c(nf_0)] = 0 \quad \text{o} \quad \pm 180° \tag{11b}$$

donde $\pm 180°$ corresponde a $c(nf_0)$ cuando es negativo. A la inversa, si una señal real tiene *simetría impar*, es decir,

$$v(-t) = -v(t) \tag{12a}$$

entonces $c(nf_0)$ es completamente *imaginaria* y

$$\arg[c(nf_0)] = \pm 90° \tag{12b}$$

lo cual se sustenta en el hecho de que $\pm j = e^{\pm j\pi/2} = e^{\pm j90°}$. La prueba de estas relaciones de simetría, se deja al lector.

Un detalle final antes de abordar un problema: cuando $v(t)$ es real, podemos escribir la ecuación (10) y reagrupar la serie exponencial en pares de fasores conjugados de la forma

$$c(nf_0)e^{j2\pi nf_0 t} + c^*(nf_0)e^{-j2\pi nf_0 t} = 2|c(nf_0)| \cos(2\pi nf_0 t + \arg[c(nf_0)])$$

de tal manera que la ecuación (6) resulta ser

$$v(t) = c(0) + \sum_{n=1}^{\infty} |2c(nf_0)| \cos(2\pi nf_0 t + \arg[c(nf_0)]) \tag{13}$$

Con este proceso hemos llegado a la serie *trigonométrica* de Fourier, y en cierta forma hemos de completar el círculo, ya que $v(t)$ queda ahora descrita más como una suma de sinusoides que de fasores conjugados. Si bien la ecuación (13) es menos versátil que la ecuación (6) y realmente carece del atractivo de la simetría que existe en la serie exponencial. Sin embargo, tendremos algunas oportunidades de usarla, así como el espectro de líneas de frecuencia positiva correspondiente.

Ejemplo 2.2 Tren de pulsos rectangulares

Como un ejemplo importante de las ideas que hemos examinado, encontremos el espectro de líneas de la forma de onda periódica de la figura 2.7, conocida como tren

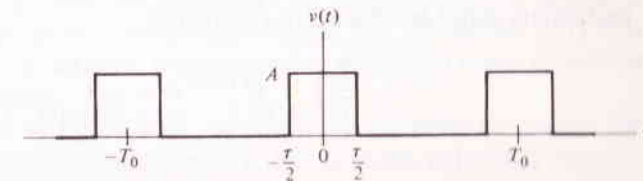



FIGURA 2.7
Tren de pulsos rectangulares

de pulsos rectangulares. Para calcular $c(nf_0)$, tomemos el intervalo de integración, en la ecuación (7), como† $[-T_0/2, T_0/2]$ y observemos que en este intervalo

$$v(t) = \begin{cases} A & |t| < \dfrac{\tau}{2} \\ 0 & |t| > \dfrac{\tau}{2} \end{cases}$$

donde τ es la duración del pulso y A es la amplitud. Nótese, en forma casual, que este modelo de señal tiene *discontinuidades* en $t = \pm\tau/2$, etc., y los valores de $v(t)$ quedan *indefinidos* dondequiera que es discontinua. Esto ilustra una de las posibles diferencias entre una señal física y su modelo matemático, ya que una señal física nunca tiene transiciones abruptas. Sin embargo, este modelo es útil si el tiempo de transición efectivo es muy corto comparado con la duración del pulso, y los valores indefinidos en los puntos de discontinuidad no tienen efecto alguno en el cálculo de $c(nf_0)$.

Continuando con el cálculo, tenemos

$$\begin{aligned} c(nf_0) &= \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} v(t)e^{-j2\pi nf_0 t}\,dt \\ &= \frac{1}{T_0}\int_{-\tau/2}^{\tau/2} Ae^{-j2\pi nf_0 t}\,dt \\ &= \frac{A}{-j2\pi nf_0 T_0}(e^{-j\pi nf_0\tau} - e^{+j\pi nf_0\tau}) \\ &= \frac{A}{\pi n}\operatorname{sen}\pi nf_0\tau \end{aligned}$$

donde nos hemos valido del hecho de que $f_0T_0 = 1$ y $e^{j\phi} - e^{-j\phi} = 2j\operatorname{sen}\phi$.

Para simplificar un poco la notación en el resultado anterior, presentamos aquí una nueva función denominada la función *senc*† y definida por

$$senc\,\lambda \triangleq \frac{\operatorname{sen}\pi\lambda}{\pi\lambda} \tag{14}$$

donde λ es la variable independiente. Esta función será de mucha importancia debido a la relación que tiene con el cálculo de promedios exponenciales y senoides; en particular, como se demuestra fácilmente

$$\frac{1}{T}\int_{-T/2}^{T/2} e^{\pm j2\pi ft}\,dt = \frac{1}{T}\int_{-T/2}^{T/2}\cos 2\pi ft\,dt = senc\,fT \tag{15}$$

† La notación $[-T_0/2, T_{0/2}]$ significa $-T_{0/2} \le t \le \tau_0/2$.

† Algunos autores emplean la llamada *función de muestreo*, Sa (λ) $\triangleq$ (sen λ)/λ; nótese que senc λ = Sa (πλ). La función *senc*, a veces se escribe *SINC*, como una incoherencia a la traducción.

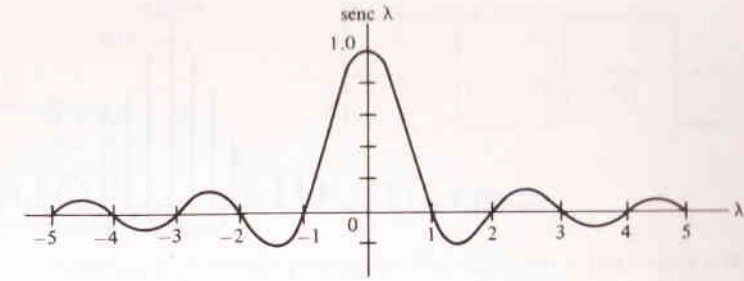



FIGURA 2.8
La función senc λ = (senc πλ)/πλ.

en la cual T es una constante arbitraria no necesariamente relacionada con f. La figura 2.8 muestra que senc λ es una función par que tiene su valor pico en $\lambda = 0$ e intersecciones con el nivel de referencia (valor cero) en todos los otros valores enteros de λ, es decir,

$$\operatorname{senc}\lambda = \begin{cases} 1 & \lambda = 0 \\ 0 & \lambda = \pm 1, \pm 2, \ldots \end{cases}$$

En la tabla C se proporcionan algunos valores numéricos de senc λ y de senc² λ.

Empleando la función senc los coeficientes de la serie para el tren de pulsos rectangulares son

$$c(nf_0) = Af_0\tau\frac{\operatorname{sen}\pi nf_0\tau}{\pi nf_0\tau} = Af_0\tau\operatorname{senc} nf_0\tau \tag{16}$$

lo cual es independiente del tiempo y estrictamente real, este último porque $v(t)$ es real y par. Por lo que el espectro de amplitud es $|c(nf_0)| = Af_0\tau\,|\operatorname{senc} nf_0\tau|$, como se muestra en la figura 2.9*a* para el caso donde $f_0\tau = 1/4$. Tales gráficas se facilitan observando la función continua $Af_0\tau|\operatorname{senc} f\tau|$ como la *envolvente* de las líneas, indicada por la curva punteada. Las líneas espectrales en $\pm 4f_0$, $\pm 8f_0$, etc., "faltan", puesto que quedan precisamente en múltiplos de $1/\tau$ donde la envolvente es igual a cero. La componente de C.D. tiene una amplitud $c(0) = Af_0\tau = A\tau/T_0$, la cual se conoce como valor promedio de $v(t)$; obsérvese la figura 2.7. Nótese que τ/T_0 es la relación del tiempo "justo" al periodo, frecuentemente designado como el *ciclo de trabajo* en electrónica de pulsos.

El espectro de fase de la figura 2.9*b*, se construye tomando en cuenta que $c(nf_0)$ es siempre real pero algunas veces negativa. Por lo tanto, el $[c(nf_0)]$ adquiere valores 0° y ±180°, según la polaridad de senc $nf_0\tau$. Los valores +180° y −180° se usan para evitar la simetría impar, aunque ello es más o menos arbitrario en tales casos.

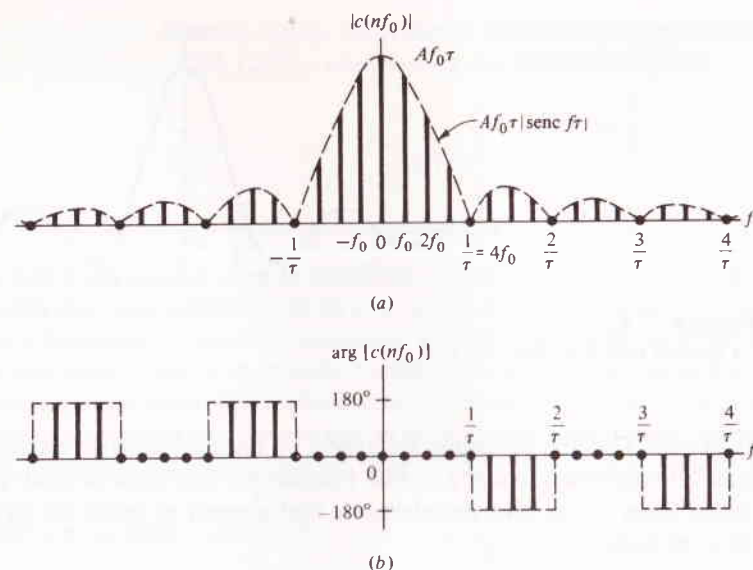



FIGURA 2.9
Espectro de un tren de pulsos rectangulares con $f_0\tau = 1/4$. *a*) Amplitud; *b*) fase.

EJERCICIO 2.6 Bosqueje el espectro de amplitud de un tren de pulsos rectangulares para cada uno de los siguientes casos $\tau = T_0/5$, $\tau = T_0/2$, $\tau = T_0$. En el último caso el tren de pulsos degenera en un valor constante durante todo el tiempo; ¿cómo se muestra éste en el espectro?

EJERCICIO 2.7 Demuestre que en la *onda cuadrada* de la figura 2.10 se tiene

$$c(nf_0) = \begin{cases} A \operatorname{senc} \dfrac{n}{2} & n = \pm 1, \pm 3, \ldots \\ 0 & n = 0, \pm 2, \pm 4, \ldots \end{cases} \tag{17}$$

(*Sugerencia*: el desarrollo mostrado a continuación, es útil para simplificar la respuesta en este problema y similares, particularmente cuando ϕ_1 o ϕ_2 son cero.)

$$e^{j\phi_1} \pm e^{j\phi_2} = [e^{j(\phi_1-\phi_2)/2} \pm e^{-j(\phi_1-\phi_2)/2}]e^{j(\phi_1+\phi_2)/2}$$
$$= \begin{cases} 2\cos\dfrac{\phi_1-\phi_2}{2}\, e^{j(\phi_1+\phi_2)/2} \\ 2j\,\operatorname{sen}\dfrac{\phi_1-\phi_2}{2}\, e^{j(\phi_1+\phi_2)/2} \end{cases} \tag{18}$$

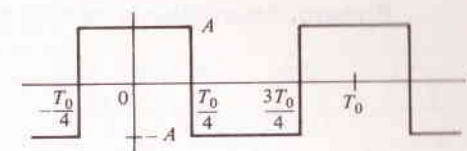



FIGURA 2.10
Onda cuadrada

El teorema de la potencia de Parseval

Este famoso teorema relaciona la potencia promedio P de una señal periódica con sus coeficientes de serie de Fourier. Para deducir la relación existente, empezamos con la definición de P, ecuación (4), escriba $|v(t)|^2 = v(t)v^*(t)$, reemplace $v^*(t)$ por su serie de Fourier e intercambie el orden de suma e integración, como sigue

$$P = \frac{1}{T_0}\int_{T_0} v(t)\overbrace{\left[\sum_{n=-\infty}^{\infty} c^*(nf_0)e^{-j2\pi nf_0t}\right]}^{v^*(t)} dt$$
$$= \sum_{n=-\infty}^{\infty} \underbrace{\left[\frac{1}{T_0}\int_{T_0} v(t)e^{-j2\pi nf_0t}\,dt\right]}_{c(nf_0)} c_v^*(nf_0)$$

Así

$$P = \sum_{n=-\infty}^{\infty} c(nf_0)c^*(nf_0) = \sum_{n=-\infty}^{\infty} |c(nf_0)|^2 \tag{19}$$

La interpretación espectral de este resultado es extraordinariamente simple, es decir, que la potencia promedio se puede encontrar elevando al cuadrado y sumando las alturas de las líneas de amplitud. Obsérvese que la ecuación (19) no incluye el espectro de fase o arg $[c(nf_0)]$, recalcando nuestro comentario previo acerca del importante papel que desempeña el espectro de amplitud en relación con el contenido de frecuencia de la señal.

Para una mejor interpretación de la ecuación (19), recordemos que la serie exponencial de Fourier desarrolla a $v(t)$ como una suma de fasores, cada uno de la forma $c(nf_0)e^{j\pi nf_0t}$. Ahora, se demuestra fácilmente que la potencia promedio de cada uno de ellos es

$$\langle |c(nf_0)e^{j2\pi nf_0t}|^2\rangle = |c(nf_0)|^2$$

Por lo tanto, el teorema de Parseval implica una *superposición de potencia promedio* en la que la potencia promedio total de $v(t)$ es la suma de las potencias promedio de sus componentes fasoriales.

Respuesta periódica de estado permanente

En la sección 2.1 encontramos la respuesta de C.A. de estado permanente de una red, expresada la entrada senoidal como una suma de fasores. Estos resultados se extienden rápidamente al caso de una señal periódica de entrada arbitraria, con el método que ahora se presenta.

Primero, desarróllese la entrada $x(t)$ como

$$x(t) = \sum_{n=-\infty}^{\infty} c_x(nf_0)e^{j2\pi nf_0t} \tag{20a}$$

donde $c_x(nf_0)$ se obtuvo de la ecuación (7). Entonces, puesto que cada uno de estos fasores produce una salida de la forma $H(nf_0)c_x(nf_0)$, exp $(j2\pi nf_0t)$, y puesto que se aplica superposición, la salida total es

$$y(t) = \sum_{n=-\infty}^{\infty} c_y(nf_0)e^{j2\pi nf_0t} \tag{20b}$$

donde

$$c_y(nf_0) = H(nf_0)c_x(nf_0) \tag{21}$$

Por lo tanto, la respuesta periódica de estado permanente es una señal periódica que tiene la misma frecuencia fundamental que la de entrada, cuyos coeficientes de serie de Fourier $c_y(nf_0)$ son iguales a los coeficientes respectivos de la señal de entrada multiplicados por la función de transferencia de la red $H(f)$ calculados en $f = nf_0$. Más aún, la potencia promedio en la señal de salida es

$$P_y = \sum_{n=-\infty}^{\infty} |c_y(nf_0)|^2 = \sum_{n=-\infty}^{\infty} |H(nf_0)|^2 |c_x(nf_0)|^2 \tag{22}$$

por aplicación del teorema de Parseval.

La interpretación del dominio de frecuencia de estos resultados se aprecia mejor convirtiendo la ecuación (21) a la forma polar, así:

$$\begin{aligned} |c_y(nf_0)| &= |H(nf_0)|\ |c_x(nf_0)| \\ \arg[c_y(nf_0)] &= \arg[c_x(nf_0)] + \arg[H(nf_0)] \end{aligned} \tag{23}$$

Poniendo esto en palabras, el espectro de amplitud de salida es igual al espectro de amplitud de entrada *veces* la relación de amplitudes de la red, mientras que el espectro de fase de salida es igual al espectro de fase de entrada *más* el corrimiento de fase de la red a las frecuencias en cuestión. La relación de fase es aditiva en lugar de multiplicativa, simplemente porque los argumentos se suman cuando las exponenciales se multiplican.

Las ecuaciones (20) a la (22) proporcionan una solución teórica al problema del análisis periódico de estado permanente. Hablando en forma práctica, sin embargo, la determinación de la *forma* real de la onda de salida $y(t)$ de la ecuación (20*b*) es un proceso tedioso —a menos que haya sólo unos pocos términos significativos en la suma—. El siguiente ejemplo cae en esta categoría y demuestra cómo la interpretación espectral ayuda a calcular el número de términos significativos.

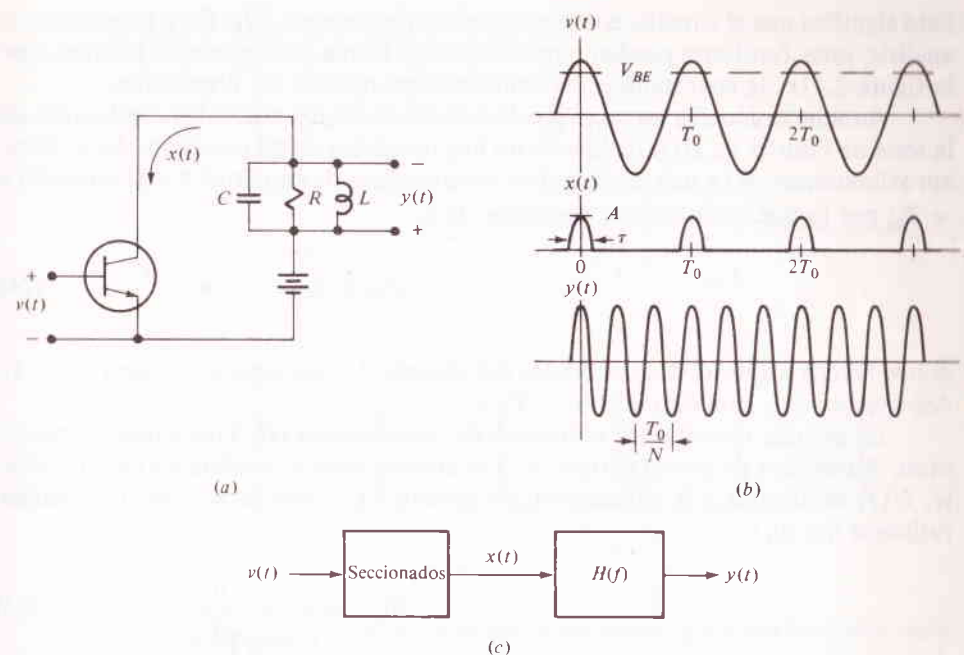



FIGURA 2.11
Multiplicador de frecuencia, *a*) Circuito; *b*) formas de onda; *c*) diagrama a cuadros.

Ejemplo 2.3 Multiplicador de frecuencia

La figura 2.11*a* es el diagrama de circuito de un multiplicador de frecuencia, un dispositivo muy usado en sistemas de comunicación. El voltaje de entrada $v(t)$ es una sinusoide a una frecuencia específica f_0 y la salida $y(t)$ se supone que es una sinusoide a un múltiplo de la frecuencia de entrada, digamos Nf_0. Generalmente $N = 2$ o 3, es decir, se tiene un *doblador* o *triplificador de frecuencia*; se obtienen factores de multiplicación mayores conectando dobladores y triplicadores en cascada (o "en tándem").

Las formas de onda mostradas en la figura 2.11*b* ilustran las operaciones esenciales, como sigue. El transistor es de silicio tipo *npn* sin polarización entre base y emisor, de tal manera que la corriente $x(t)$ puede ser pasada por alto a menos que $v(t)$ exceda el valor de la caída de voltaje entre base y emisor V_{BE}. Así, si la amplitud de $v(t)$ es sólo ligeramente mayor que V_{BE}, como se indica, $x(t)$ estará formada por pulsos cortos cuyo periodo será $T_0 = 1/f_0$. De nuestro estudio previo sabemos que tal forma de onda contendrá en general *todas las armónicas* de f_0, y que la función del circuito paralelo RLC† es seleccionar la n-ésima armónica y rechazar al resto.

† Como norma, la resistencia R representa las pérdidas en L y cualquier carga acoplada y no un elemento de circuito formal.

Esto significa que el circuito está *sintonizado* o *resonante a* Nf_0. Para propósitos de análisis, estas funciones pueden representarse en forma de diagrama a bloques, por la figura 2.11*c*, la cual tomaremos como nuestro modelo del dispositivo.

Aunque se conciba que uno puede calcular en forma exacta los coeficientes de la serie de Fourier de $x(t)$, realmente no hay necesidad de tal precisión. Antes bien, aproximaremos $x(t)$ a un tren de pulsos *rectangulares* de amplitud A y de duración $\tau \ll T_0$; por lo que, para valores pequeños de n,

$$c_x(nf_0) \approx Af_0\tau \qquad |n| \ll \frac{T_0}{\tau} \tag{24}$$

donde hemos empleado los resultados del ejemplo 2.2 con senc $\lambda \approx 1$ para $|\lambda| \ll 1$. Asimismo, se ha supuesto que $N \ll T_0/\tau$.

En seguida necesitamos la función de transferencia $H(f)$ del circuito sintonizado. Advertidos de que su entrada $x(t)$ es una corriente y la salida $y(t)$ es un voltaje, $H(f)$ es idéntica a la impedancia del circuito $Z(j\omega)$ con $\omega = j2\pi f$. Un análisis rutinario nos da †

$$H(f) = \frac{R}{1 + jQ\left(\dfrac{f^2 - f_r^2}{ff_r}\right)} \tag{25}$$

donde

$$f_r \triangleq \frac{1}{2\pi\sqrt{LC}} \qquad Q \triangleq \frac{R}{2\pi f_r L} = R\sqrt{\frac{C}{L}}$$

las cuales son la frecuencia resonante y el factor de calidad, respectivamente. Se ha supuesto en la ecuación (25) que $Q > ½$, por lo que el circuito realmente es resonante; aunque de hecho, la presente aplicación requiere que $Q \gg 1$. La correspondiente relación de amplitudes $H(f)$ está en la gráfica de la figura 2.12*a* directamente arriba del espectro de amplitud de entrada $|c_x(nf_0)|$, figura 2.12*b*. Se han omitido, por conveniencia, las frecuencias negativas, puesto que sabemos que ambas funciones tienen simetría par.

Si recordamos $|c_y(nf_0)| = |H(nf_0)|\ |c_x(nf_0)|$, la figura 2.12 sugiere que todas las armónicas de $y(t)$ con excepción de Nf_0 pueden considerarse de amplitud despreciable si

$$f_r = Nf_0 \qquad \text{y} \qquad Q \gg N/2$$

la condición en Q para asegurar que $[1 + (½Q)]f_r < (N + 1)f_0$ y $[1 - (½Q)f_r] > (N - 1)f_0$ por lo que el circuito sintonizado permitirá sólo el "paso" de Nf_0. Por lo

† Véase, por ejemplo, Close (1966, Cap. 6).

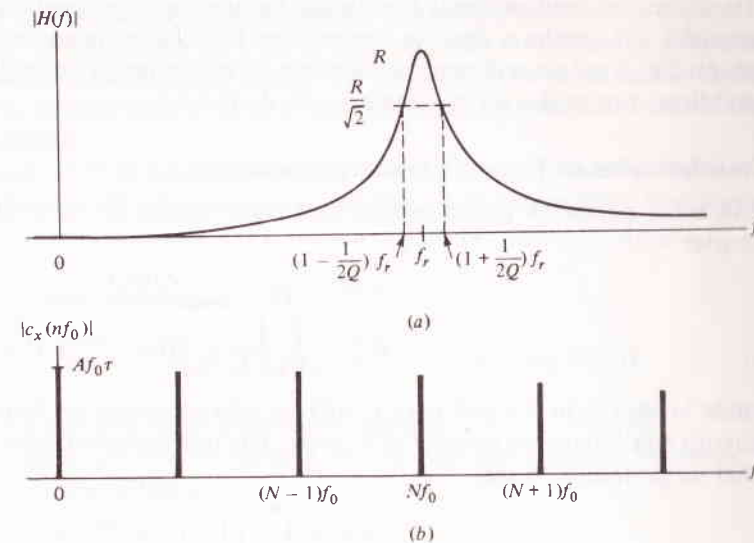



FIGURA 2.12
a) Relación de amplitudes de un circuito sintonizado, $Q = 10$; *b*) Espectro de amplitud de la señal de entrada.

tanto, el espectro de amplitud de salida consiste esencialmente de dos términos, $c_y(\pm Nf_0)$, y

$$y(t) \approx H(Nf_0)c_x(Nf_0)e^{j2\pi Nf_0t} + H(-Nf_0)c_x(-Nf_0)e^{-j2\pi Nf_0t}$$
$$= 2RAf_0\tau \cos 2\pi Nf_0 t$$

donde hemos empleado las ecuaciones (24) y (25) y convertido el resultado a la forma senoidal.

2.3 SEÑALES NO PERIÓDICAS Y TRANSFORMADAS DE FOURIER

Hemos descrito a una señal periódica como aquella cuyas características se repiten durante un largo intervalo de tiempo, teóricamente infinito. Ahora consideremos señales no periódicas cuyos efectos se concentran en un breve periodo. Tales señales pueden ser *estrictamente limitadas en tiempo*, es decir que $v(t)$ es idénticamente cero fuera del intervalo considerado, o *asintóticamente limitada en tiempo*, es decir que $v(t) \to 0$ cuando $t \to \pm\infty$. En uno u otro caso, se supone que la *energía* total de la señal está bien definida, y medida en la misma forma normalizada como lo fue la potencia en la sección anterior, o sea,

$$E \triangleq \int_{-\infty}^{\infty} |v(t)|^2\, dt \tag{1}$$

Esta definición implica que si E es finita, tanto el valor promedio como la potencia promedio son iguales a cero, se contınúa un análisis en forma más completa en la sección 2.6. Aquí describiremos el dominio de la frecuencia de señales de energía no periódicas, por medio de la transformada de Fourier.

Transformadas de Fourier y espectros continuos

Uña señal periódica puede representarse por medio de su serie exponencial de Fourier

$$v(t) = \sum_{n=\infty}^{\infty} \overbrace{\left[\frac{1}{T_0}\int_{T_0} v(t)e^{-j2\pi nf_0t}\,dt\right]}^{c_v(nf_0)} e^{j2\pi nf_0t} \tag{2}$$

donde la expresión integral para $c_v(nf_0)$ ha sido transcrita en forma completa. De acuerdo con el *teorema integral de Fourier*, hay una representación similar para una señal *no periódica*, o sea,

$$v(t) = \int_{-\infty}^{\infty} \underbrace{\left[\int_{-\infty}^{\infty} v(t)e^{-j2\pi ft}\,dt\right]}_{V(f)} e^{j2\pi ft}\,df \tag{3}$$

El término entre paréntesis rectangulares es la *transformada de Fourier* de $v(t)$, simbolizada por $V(f)$ o [$v(t)$] definida como

$$V(f) = \mathscr{F}[v(t)] \triangleq \int_{-\infty}^{\infty} v(t)e^{-j2\pi ft}\,dt \tag{4}$$

la cual es una integración sobre todo el tiempo. El teorema (3) establece que $v(t)$ se puede encontrar por medio de la *transformada de Fourier inversa* de $V(f)$,

$$v(t) = \mathscr{F}^{-1}[V(f)] \triangleq \int_{-\infty}^{\infty} V(f)e^{j2\pi ft}\,df \tag{5}$$

la cual es una integración sobre todos los valores de frecuencia.

A las ecuaciones (4) y (5) se les designa a menudo como integrales de Fourier o como par de transformadas de Fourier † y a primera vista, parecen ser un circulo cerrado de operaciones. En un problema dado, se conoce generalmente ya sea $V(f)$ o $v(t)$, pero no ambas. Si $V(f)$ es conocida, podemos encontrar $v(t)$ recurriendo a la transformada inversa (5), y viceversa, cuando queremos encontrar $V(f)$ a partir de $v(t)$.

Si regresamos a la gráfica del dominio de la frecuencia y comparamos las ecuaciones (2) y (3) vemos que $V(f)$ juega el mismo papel para señales no periódicas que el que juega $c_v(nf_0)$ para señales periódicas. Así, $V(f)$ es el *espectro* de la señal no periódica $v(t)$. Pero $V(f)$ es una función continua definida para todos los valores de f, mientras que $c_v(nf_0)$ está definida sólo para frecuencias discretas. Por lo

† Se aplican definiciones algo diferentes cuando se emplea $\omega = 2\pi f$ como la variable independiente del dominio de la frecuencia.

tanto, una señal no periódica tendrá un *espectro continuo* y no sólo un espectro de líneas. Comparar las ecuaciones (2) y (3) ayuda a explicar la diferencia: en el caso periódico regresamos al dominio del tiempo por medio de la *suma* de fasores de frecuencias discretas, mientras que en el caso no periódico *integramos* una función continua de la frecuencia.

Como $c_v(nf_0)$, $V(f)$ es generalmente una función compleja donde $|V(f)|$ es el espectro de amplitud y [$V(f)$] es el espectro de fase. A continuación relacionamos otras propiedades importantes de $V(f)$, paralelas a las propiedades $c_v(nf_0)$,

1 Si $v(t)$ es *real*, entonces $V(-f) = V^*(f)$ y

$$|V(-f)| = |V(f)| \qquad \arg[V(-f)] = -\arg[V(f)] \tag{6}$$

Por lo que el espectro tiene simetría hermitiana.

2 Si $v(t)$ tiene *simetría en tiempo* ya sea por par o impar, entonces la ecuación (4) se simplifica a

$$V(f) = \begin{cases} 2\displaystyle\int_0^{\infty} v(t)\cos\omega t\,dt & v(t)\text{ par} \\ -j2\displaystyle\int_0^{\infty} v(t)\operatorname{sen}\omega t\,dt & v(t)\text{ impar} \end{cases} \tag{7}$$

donde hemos escrito ω en lugar de $2\pi f$ por conveniencia de anotación; una práctica que será usada frecuentemente en lo sucesivo. De la ecuación (7) se tiene que si $v(t)$ es también real, entonces $V(f)$ es puramente real o imaginaria, respectivamente.

3 El valor de $V(f)$ en $f = 0$ es igual al *área neta* de $v(t)$, es decir,

$$V(0) = \int_{-\infty}^{\infty} v(t)\,dt \tag{8}$$

lo cual se compara con el caso periódico donde $c_v(0)$ es igual al valor promedio de $v(t)$.

EJERCICIO 2.8 Las integrales de la forma general $\int_{-T}^{T} w(t)dt$ se simplifican cuando el integrado es simétrico, específicamente para cualquier constante T.

$$\int_{-T}^{T} w(t)\,dt = \begin{cases} 2\displaystyle\int_0^{T} w(t)\,dt & \text{si } w(-t) = w(t) \quad (9a) \\ 0 & \text{si } w(-t) = -w(t) \quad (9b) \end{cases}$$

Emplee la ecuación (9) para deducir la ecuación (7) de la ecuación (4).

FIGURA 2.13
El pulso rectangular $v(t) = A\Pi(t/\tau)$.

Ejemplo 2.4 Pulso rectangular

En el ejemplo 2.2 de la sección 2.2 encontramos el espectro de líneas de un tren de pulsos rectangulares. Ahora consideraremos el espectro continuo del pulso rectangular *sencillo* mostrado en la figura 2.13. Este es un modelo de señal tan común que merece un símbolo para identificarlo. Por lo tanto, usemos el siguiente:

$$\Pi\left(\frac{t}{\tau}\right) \triangleq \begin{cases} 1 & |t| < \dfrac{\tau}{2} \\ 0 & |t| > \dfrac{\tau}{2} \end{cases} \tag{10}$$

el cual vale para una *función rectangular* con altura o amplitud unitaria, con ancho o duración τ y centro en $t = 0$. Así, $v(t) = A\Pi(t/\tau)$ en la figura 2.13.

Puesto que $v(t)$ tiene simetría par, su transformada de Fourier es

$$\begin{aligned} V(f) &= 2\int_0^\infty v(t)\cos\omega t\, dt \\ &= 2\int_0^{\tau/2} A\cos\omega t\, dt = \frac{2A}{\omega}\operatorname{sen}\frac{\omega\tau}{2} \\ &= A\tau \operatorname{senc} f\tau \end{aligned} \tag{11}$$

y $V(0) = A\tau$ la cual claramente es igual al área del pulso. El espectro correspondiente está representado en la gráfica de la figura 2.14 que debe compararse con la 2.7 para ilustrar nuestra discusión de los espectros de líneas y continuos.

Observando $|V(f)|$ es aparente que la porción significativa del espectro está en el intervalo $|f| < 1/\tau$ puesto que $|V(f)| \ll V(0)$ para $|f| > 1/\tau$. Podemos, por lo tanto, tomar $1/\tau$ como medida del "ancho" espectral. Ahora, si la duración del pulso se reduce (τ pequeña), el ancho de la frecuencia aumenta, mientras que si se aumenta la duración se reduce el ancho espectral. Así, pulsos cortos tienen espectros anchos, pulsos largos tienen espectros angostos. Este fenómeno se conoce como *anchura recíproca* y es una propiedad general de todas las señales, pulsadas o no, porque los componentes de alta frecuencia son causadas por variaciones rápidas en tiempo,

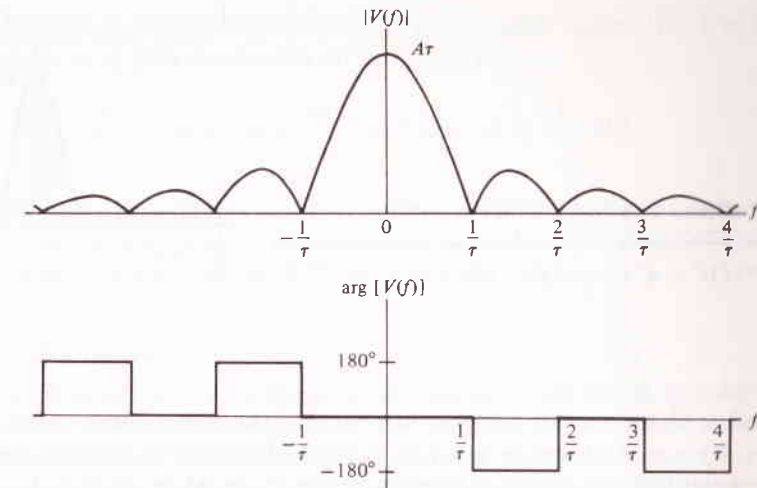



FIGURA 2.14
Espectro de un pulso rectangular, $V(f) = A\tau \operatorname{senc} f\tau$.

mientras que en las variaciones más suaves o lentas en el tiempo se requiere relativamente poco contenido de alta frecuencia.

Ejemplo 2.5 Pulso exponencial

Considere una función exponencial amortiguada de la forma $v(t) = Ae^{-t/T}$, $t > 0$. Para asegurar que la energía es finita, apoyamos especificando que $v(t) = 0$ para $t < 0$. Introduciendo la notación *funcional de paso unitario*

$$u(t) \triangleq \begin{cases} 1 & t > 0 \\ 0 & t < 0 \end{cases} \tag{12}$$

podemos escribir

$$v(t) = Ae^{-t/T}u(t)$$

a lo cual se le designará como pulso exponencial.

La integración de $V(f)$ es un problema simple, resultando

$$V(f) = \int_0^\infty Ae^{-t/T}e^{-j\omega t}\, dt = \frac{AT}{1 + j2\pi fT}$$

donde la amplitud y el espectro de fase son

$$|V(f)| = \frac{AT}{\sqrt{1 + (2\pi fT)^2}} \qquad \arg[V(f)] = -\arctan 2\pi fT$$

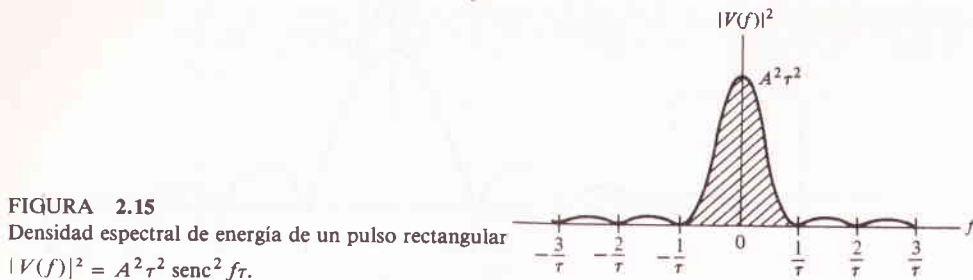



FIGURA 2.15
Densidad espectral de energía de un pulso rectangular $|V(f)|^2 = A^2\tau^2 \operatorname{senc}^2 f\tau$.

Así, a diferencia del ejemplo anterior, la fase es una curva suave entre $+90°(f = -\infty)$ y $-90°(f = +\infty)$. Los lectores observadores probablemente encontrarán semejanza entre $V(f)$ y la función de transferencia de un filtro pasabajas RC, en el ejemplo 2.1. La semejanza no es accidental, y explicaremos por qué en la sección 2.5

Teorema de la energía de Rayleigh

Este teorema, análogo al de la potencia de Parseval, relaciona la energía total E de una señal con su espectro de amplitud. En realidad, el teorema es un caso especial de una interesante relación integral

$$\int_{-\infty}^{\infty} v(t)w^*(t)\,dt = \int_{-\infty}^{\infty} V(f)W^*(f)\,d_f \tag{13}$$

donde $V(f) = \mathscr{F}[v(t)]$ y $W(f) = \mathscr{F}[w(t)]$. Para comprobar la ecuación (13) se sigue el mismo método empleado para comprobar el teorema de Parseval, ecuación (19), de la sección 2.2 El teorema de Rayleigh se obtiene haciendo $w(t) = v(t)$, así, el primer miembro de la ecuación (13) es la energía de $v(t)$ —por la definición (1)— y, por lo tanto

$$E = \int_{-\infty}^{\infty} V(f)V^*(f)\,df = \int_{-\infty}^{\infty} |V(f)|^2\,df \tag{14}$$

Por lo que, integrando el cuadrado del espectro de amplitud $|V(f)|^2$ sobre todos los valores de la frecuencia, ello nos da la energía total.

El valor de la ecuación (14) no depende mucho del cálculo de E, puesto que la integración en el dominio del tiempo de $|v(t)|^2$ a menudo es más fácil. Es más, ello implica que $|V(f)|^2$ da la distribución de energía en el domino de la frecuencia y, por lo tanto, se le puede designar como la *densidad espectral de energía*. Por esto se debe entender que la energía en cualquier banda de frecuencia diferencial $f \pm df/2$ es igual a $|V(f)|^2\,df$. Esa interpretación, a su vez, presta apoyo cuantitativo a la noción del ancho espectral, en el sentido de que mucha de la energía de una señal dada debe estar contenida en el intervalo de frecuencias que vienen a ser el ancho espectral.

A modo de ilustración, la figura 2.15 muestra la densidad espectral de energía

de un pulso rectangular cuyo ancho espectral se fijó previamente como $|f| < 1/\tau$. La energía en esa banda es el área sombreada en la figura, es decir,

$$\int_{-1/\tau}^{1/\tau} |V(f)|^2\,df = \int_{-1/\tau}^{1/\tau} (A\tau)^2 \operatorname{senc}^2 f\tau\,df = 0.92A^2\tau$$

cuya evaluación requiere de métodos numéricos. Pero la energía total de la señal es $E = \int_{-\infty}^{\infty} |v(t)|^2\,dt = A^2\tau$ (¡encontrada por observación!) de tal manera que el ancho espectral seguro incluye más del 90% de la energía total.

Teoremas de transformadas

En seguida se enuncian muchos otros teoremas asociados con las transformadas de Fourier. Se incluyen no sólo como ejercicios de práctica, sino también por dos buenas razones. Primera, los teoremas son de gran valor para la interpretación de espectros, porque expresan relaciones entre operaciones del dominio del tiempo y del dominio de la frecuencia. Segunda, podemos construir una lista extensa de pares de transformadas aplicando los teoremas a pares conocidos —y esta lista nos será útil conforme busquemos nuevos modelos de señales.

Para establecer los teoremas, indicamos una señal y su transformada (o espectro) con letras mayúsculas y minúsculas, por ejemplo, $V(f) = \mathscr{F}[v(t)]$ y $v(t) = \mathscr{F}^{-1}[V(f)]$. Esto también se denota en forma más compacta por $v(t) \longrightarrow V(f)$. La tabla A al final del libro contiene los teoremas y pares de transformadas citados aquí y algunos otros.

Linealidad (superposición)

Para las constantes α y β

$$\alpha v(t) + \beta w(t) \leftrightarrow \alpha V(f) + \beta W(f) \tag{15}$$

Este teorema simplemente establece que las combinaciones lineales en el dominio del tiempo resultan ser combinaciones lineales en el dominio de la frecuencia. Aunque la comprobación del teorema es trivial, su importancia no puede ser reenfantizada. Desde un punto de vista práctico de ecuación (15) facilita mucho el análisis espectral cuando la señal en cuestión es una combinación lineal de funciones cuyos espectros individuales son conocidos. Desde un punto de vista teórico, subraya la facilidad de aplicación de la transformada de Fourier al estudio de los sistemas lineales.

Retardo en el tiempo

Si una señal $v(t)$ se retrasa en el tiempo en t_d segundos, originando la nueva señal $v(t - t_d)$ el espectro se modifica por un corrimiento de fase lineal o pendiente $-2\pi t_d$, o sea,

$$v(t - t_d) \leftrightarrow V(f)e^{-j\omega t_d} \tag{16}$$

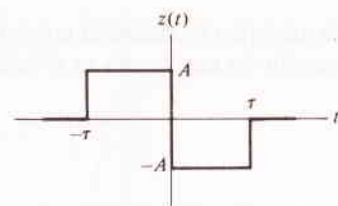



FIGURA 2.16

La traslación de una señal en el tiempo, en esta forma, cambia la fase espectral pero no la amplitud. Nótese que si t_d es un número negativo, la señal *avanza* en el tiempo y el corrimiento de fase tiene pendiente *positiva*. Puesto que un avance en el tiempo es una imposibilidad física, concluimos que en el procesamiento real de la señal, la fase espectral tendrá pendiente negativa, aunque no necesariamente lineal.

La comprobación de este teorema se completa con un cambio de variable $\lambda = t - t_d$ en la integral transformada. (Obsérvese que el tiempo es en realidad una variable *fantasma* en la transformada directa, así como la frecuencia es una variable *fantasma* en la transformada inversa.) La técnica del cambio de variable es indispensable para la prueba de muchos teoremas de transformadas; se demuestra aquí que

$$\begin{aligned}\mathscr{F}[v(t-t_d)] &= \int_{-\infty}^{\infty} v(t-t_d)e^{-j\omega t}\,dt \\ &= \int_{-\infty}^{\infty} v(\lambda)e^{-j\omega(\lambda+t_d)}\,d\lambda \\ &= \left[\int_{-\infty}^{\infty} v(\lambda)e^{-j\omega\lambda}\,d\lambda\right]e^{-j\omega t_d}\end{aligned}$$

La integral entre paréntesis rectangulares es precisamente $V(f)$, así, $\mathscr{F}[v(t-t_d)] = V(f)e^{-j\omega t_d}$.

EJERCICIO 2.9 La señal de la figura 2.16 se puede escribir como

$$z(t) = A\Pi\left(\frac{t+t_d}{\tau}\right) - A\Pi\left(\frac{t-t_d}{\tau}\right) \qquad t_d = \frac{\tau}{2} \tag{17a}$$

Aplicamos los teoremas de linealidad y retardo en tiempo a los resultados del ejemplo 2.4 para obtener

$$Z(f) = j2A\tau \operatorname{senc} f\tau \operatorname{sen} \pi f\tau \tag{17b}$$

Luego bosquejamos el espectro de amplitud y comparamos con la figura 2.14.

Cambio de escala

El retardo en el tiempo es equivalente a la traslación del origen del tiempo. Otra operación geométrica es el cambio de escala, en la cual el eje del tiempo se expande, se comprime o se invierte. Así, $v(at)$ es una versión comprimida de $v(t)$ siendo a un número positivo y mayor que 1. De manera similar, si a es negativo y menor que 1, $v(at)$ es la imagen extendida de $v(t)$ inversa en el tiempo. Estas operaciones pueden ocurrir con señales grabadas cuando se reproducen, por ejemplo.

El teorema del cambio de escala dice que

$$v(at) \leftrightarrow \frac{1}{|a|} V\left(\frac{f}{a}\right) \tag{18}$$

lo cual expresa formalmente la propiedad de la anchura recíproca encontrada en el ejemplo 2.4; porque si la señal se comprime en tiempo por medio del factor a, su espectro se extiende en frecuencia por medio de $1/a$ e inversamente. El teorema se prueba por cambio de variables, considerando los valores positivos y negativos de a en forma separada.

Dualidad

El concepto poderoso de la dualidad es bien conocido en el análisis de circuitos. En el análisis espectral también existe dualidad entre los dominios del tiempo y la frecuencia, la cual se origina de la semejanza con las integrales transformadas de Fourier. El teorema de la dualidad dice que si

$$v(t) \leftrightarrow V(f)$$

entonces la transformación de la función del *tiempo* $V(t)$ es

$$\mathscr{F}[V(t)] = v(-f) \tag{19}$$

como se comprueba intercambiando t y f en las integrales transformadas de Fourier.

En la forma de la ecuación (19) este teorema es muy abstracto, y es difícil visualizar su uso para la generación de nuevos pares de transformadas. El siguiente ejemplo debe ayudar a aclarar el procedimiento.

Ejemplo 2.6 Pulso senc

Considérese la señal $z(t) = A \operatorname{senc} 2Wt$, o sea, una función senc en el tiempo. (Aunque la idea de un pulso senc pueda parecer extraña al principio, juega un papel importante en el estudio de la transmisión digital de datos.) Recordando el par de transformadas del ejemplo 2.4, $A\Pi(t/\tau) \longleftrightarrow A\tau \operatorname{sen} cf\tau$, aplicamos la dualidad escribiendo $z(t)$ en la forma

$$z(t) = V(t) = \frac{A}{2W} 2W \operatorname{senc} t2W$$

así

$$Z(f) = v(-f) = \frac{A}{2W}\Pi\left(\frac{-f}{2W}\right)$$

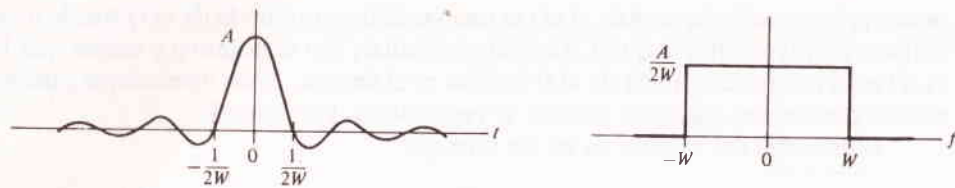



FIGURA 2.17
Un pulso senc y su espectro de banda limitada.

Como la función rectangular tiene simetría par, $\Pi(-f/2W) = \Pi(f/2W)$, y hemos deducido el nuevo par de transformadas

$$A \operatorname{senc} 2Wt \leftrightarrow \frac{A}{2W} \Pi\left(\frac{f}{2W}\right) \qquad (20)$$

lo cual se muestra en la figura 2.17. Como se puede ver, el espectro de un pulso senc tiene definido en forma clara el ancho espectral W. De hecho, el espectro es cero para $|f| > W$, y se dice que la señal es de *banda limitada* en W. Éste es nuestro primer encuentro con una señal que tiene una banda limitada estrictamente, en frecuencia. Nótese que la señal en sí misma está limitada asintóticamente sólo en tiempo.

Traslación de frecuencia (modulación)

La dualidad puede emplearse para generar teoremas de transformadas así como pares de transformadas. Por ejemplo, una dualidad del teorema del retardo en el tiempo (16) es

$$v(t)e^{j\omega_c t} \leftrightarrow V(f - f_c) \qquad \omega_c = 2\pi f_c \qquad (21)$$

Designamos a esto como *traslación de frecuencia* o *modulación compleja*, puesto que al multiplicar una función del tiempo por $e^{j\omega_c t}$ da lugar a que en su espectro haya una traslación en frecuencia de $+f_c$.

Para ver los efectos de la traslación de frecuencia, consideremos que $v(t)$ tiene el espectro de banda limitada de la figura 2.18a, donde la amplitud y fase se trazan

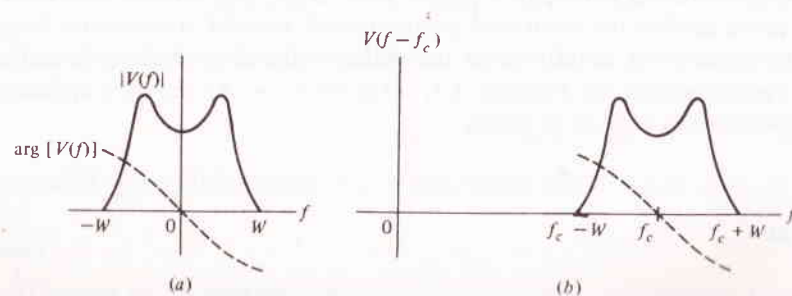



FIGURA 2.18

sobre los mismos ejes empleando líneas continuas y punteadas respectivamente. La observación del espectro trasladado $V(f - f_c)$ en la figura 2.18b revela lo siguiente:

1 Las componentes significativas están concentradas alrededor de la frecuencia f_c.
2 Aunque $V(f)$ es de banda limitada en W, $V(f - f_c)$ tiene un ancho espectral de $2W$. Por lo tanto, con la traslación se ha doblado el ancho espectral. Dicho de otra manera, la porción de frecuencia negativa de $V(f)$ aparece ahora dentro de las frecuencias positivas.
3 $V(f - f_c)$ no es hermitiana pero tiene simetría con respecto al origen trasladado a $f = f_c$.

Estas consideraciones pueden parecer algo académicas en vista de que $v(t)e^{j\omega_c t}$ no es una función real del tiempo y no puede ocurrir como una señal de comunicación. No obstante, son comunes las señales de la forma $v(t) \cos(\omega_c t + \theta)$ de hecho, son éstas la base de la modulación de portadora y por extensión directa de la ecuación (21) se tiene el siguiente *teorema de la modulación*:

$$v(t) \cos(\omega_c t + \theta) \leftrightarrow \frac{e^{j\theta}}{2} V(f - f_c) + \frac{e^{-j\theta}}{2} V(f + f_c) \qquad (22)$$

En palabras: multiplicando una señal por una sinusoide, se traslada su espectro en frecuencia, *arriba y abajo* de f_c. También se aplican aquí todos los comentarios acerca de la modulación compleja. Además, el espectro resultante es hermitiano, lo cual se cumple si $v(t) \cos(\omega_c t + \theta)$ es una función real del tiempo. El teorema se comprueba fácilmente con la ayuda del teorema de Euler y de la ecuación (21).

Ejemplo 2.7 Pulso de RF

Considérese la sinusoide de duración finita de la figura 2.19a, considerada como un pulso de RF cuando f_c cae en la banda de radiofrecuencia. Puesto que

$$z(t) = A\Pi\left(\frac{t}{\tau}\right) \cos \omega_c t \qquad (23a)$$

tenemos inmediatamente

$$Z(f) = \frac{A}{2} \operatorname{senc}(f - f_c)\tau + \frac{A}{2} \operatorname{senc}(f + f_c)\tau \qquad (23b)$$

haciendo $v(t) = A\Pi(t/\tau)$ y $V(f) = A\tau \operatorname{senc} f\tau$ en la ecuación (22). El espectro de amplitud resultante se bosqueja en la figura 2.19b para el caso en que $f_c \gg 1/\tau$ y así las dos funciones senc trasladadas tienen traslapos despreciables.

Como ésta es una sinusoide de duración finita, su espectro es continuo y contiene otras frecuencias además de $f = \pm f_c$. Estas otras frecuencias se originan por el hecho de que $z(t) = 0$ para $|t| > \tau/2$, y el valor más pequeño de τ da lugar a la mayor anchura o extensión espectral alrededor de $\pm f_c$ —(anchura recíproca, nueva-

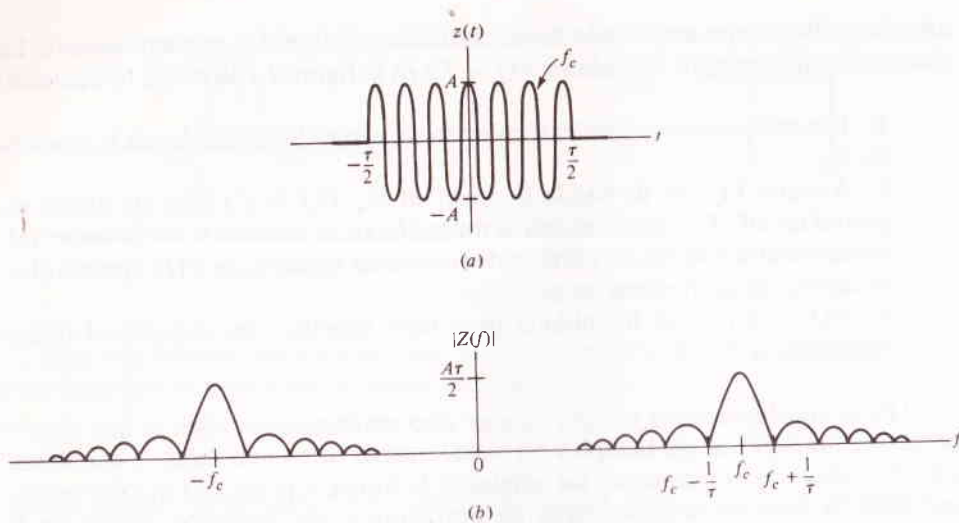



FIGURA 2.19
Un pulso de R.F. y su espectro de amplitud, $f_c \gg 1/\tau$.

mente). Por otro lado, si se tratara de una sinusoide de duración *infinita*, la representación del dominio de la frecuencia sería un espectro *de líneas* de dos lados que contuviera sólo las frecuencias discretas $\pm f_c$.

Diferenciación e integración

Ciertas técnicas de procesamiento incluyen la diferenciación o la integración de una señal. En los teoremas siguientes, se indican los efectos de estas operaciones en el dominio de la frecuencia. Sin embargo, antes una advertencia: los teoremas no se deben emplear antes de estar seguros de que la señal diferenciada o integrada se puede analizar a base de transformadas de Fourier, es deci:, que tiene energía bien definida. Además, el hecho de que $v(t)$ tiene energía finita no es una garantía como para asegurar que se puede derivar o integrar.

Para deducir el teorema de la diferenciación, sustituimos $v(t)$ por la integral transformada inversa e intercambiamos el orden de operaciones como sigue:

$$\frac{d}{dt}v(t) = \frac{d}{dt}\left[\int_{-\infty}^{\infty} V(f)e^{j2\pi ft}\,df\right]$$

$$= \int_{-\infty}^{\infty} V(f)\left(\frac{d}{dt}e^{j2\pi ft}\right)df$$

$$= \int_{-\infty}^{\infty} \underbrace{[j2\pi fV(f)]}_{\mathscr{F}\left[\frac{dv(t)}{dt}\right]}e^{j2\pi ft}\,df$$

Al referirnos al teorema integral de Fourier (3), éste revela que el término entre paréntesis rectangulares debe ser $[dv(t)/dt]$, así

$$\frac{d}{dt}v(t) \leftrightarrow j2\pi fV(f)$$

y por iteración

$$\frac{d^n}{dt^n}v(t) \leftrightarrow (j2\pi f)^n V(f) \tag{24}$$

el cual es el *teorema de la diferenciación*.

Ahora supóngase que generamos otra función del tiempo a partir de $v(t)$ al desarrollar la operación $\int_{-\infty}^{t} v(\lambda)d\lambda$, donde se requiere la variable muda λ, ya que la variable independiente t es el límite superior de integración. El *teorema de la integración* dice que

$$\int_{-\infty}^{t} v(\lambda)\,d\lambda \leftrightarrow \frac{1}{j2\pi f}V(f) \tag{25}$$

cuya demostración incluye el mismo método usado antes. También se puede generalizar la ecuación (25) a una integración múltiple, pero la notación resulta engorrosa.

Observando estos teoremas, podemos expresar que la diferenciación aumenta los componentes de alta frecuencia de una señal, en tanto que la integración los suprime. La interpretación espectral está así conforme con el punto de vista del dominio del tiempo, en el sentido de que la diferenciación acentúa las variaciones en el tiempo mientras que la integración las suaviza.

Ejemplo 2.8 Pulso triangular

Para ilustrar el teorema de la integración —y obtener aún otro par transformado útil— integremos la señal $z(t)$ de la figura 2.16 y dividámosla por la constante τ. Esto nos da

$$w(t) = \frac{1}{\tau}\int_{-\infty}^{t} z(\lambda)\,d\lambda = \begin{cases} A\left(1 - \dfrac{|t|}{\tau}\right) & |t| < \tau \\ 0 & |t| > \tau \end{cases}$$

una forma de pulso *triangular*, que se muestra en la figura 2.20*a*. (Que el lector compruebe este resultado empleando la interpretación gráfica de la integración). Entonces, tomando $Z(f)$ de la ecuación (17*b*), tenemos

$$W(f) = \frac{1}{\tau}\frac{1}{j2\pi f}Z(f) = \frac{j2A\tau\,\mathrm{senc}\,f\tau\,\mathrm{sen}\,\pi f\tau}{j2\pi f\tau} = A\tau\,\mathrm{senc}^2\,f\tau$$

como en la gráfica de la figura 2.20*b*. Al comparar este espectro con el de la figura 2.14 se ve que el pulso triangular tiene menos contenido de alta frecuencia que un pulso rectangular con amplitud A y duración τ, aun cuando ambos tengan la misma

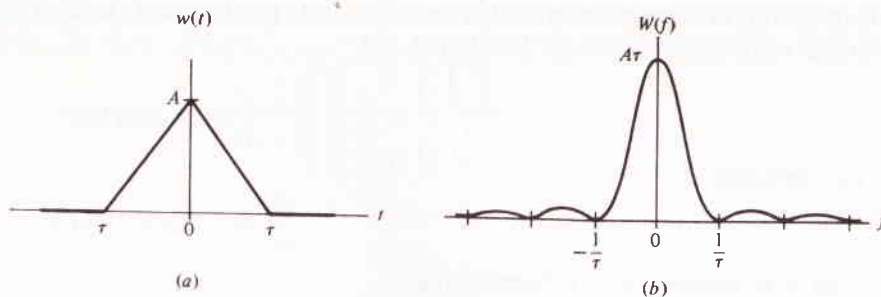



FIGURA 2.20
Un pulso triangular y su espectro.

área. La diferencia procede de que el pulso triangular se extiende sobre 2τ segundos y no tiene las marcadas variaciones en tiempo que la forma *rectangular*.

Este par transformado se puede escribir en forma más compacta por medio de la definición de la *función triangular*

$$\Lambda\left(\frac{t}{\tau}\right) \triangleq \begin{cases} 1 - \dfrac{|t|}{\tau} & |t| < \tau \\ 0 & |t| > \tau \end{cases} \tag{26}$$

Entonces y

$$A\Lambda\left(\frac{t}{\tau}\right) \leftrightarrow A\tau \operatorname{senc}^2 f\tau \tag{27}$$

Resulta así que las funciones triangulares pueden ser generadas de funciones rectangulares por medio de otra operación matemática, o sea, la convolución. Y la convolución es el principal material de nuestra agenda, por lo que pronto echaremos otro vistazo a este ejemplo.

2.4 CONVOLUCIÓN E IMPULSOS

La operación matemática conocida como convolución posee un alta clasificación entre la herramienta analítica usada por los ingenieros en comunicaciones. Por una razón, es un buen modelo de los procesos físicos que existen en un sistema lineal; por otra parte, nos ayuda a comprender las relaciones entre los dominios del tiempo y de la frecuencia. En ambos casos, va de la mano con esa curiosa ficción de ingeniería conocida como el impulso o función delta. Esta sección trata estos conceptos en cuanto a sus relaciones con las señales. Su aplicación a sistemas lineales se tratará en la siguiente sección.

Integral de convolución

La convolución de dos funciones de la misma variable, digamos $v(t)$ y $w(t)$, se define por

$$v * w(t) \triangleq \int_{-\infty}^{\infty} v(\lambda)w(t-\lambda)\, d\lambda \tag{1}$$

donde $v * w(t)$ vale solamente para la operación anotada en el segundo miembro de la ecuación (1) y el asterisco (*) no tiene nada que ver con complejos conjugados. La ecuación (1) es la *integral de convolución*, que a menudo se denota por $v * w$ cuando la variable independiente es no ambigua. En otras ocasiones la notación $[v(t)]*[w(t)]$ se hace necesaria para aclarar. Obsérvese que aquí la variable independiente es t, así como en las funciones que van a ser tratadas con este procedimiento; la integración también se lleva a cabo con respecto a una variable muda (tal como λ) y t es una constante en cuanto a la integración se refiere.

El cálculo de $v * w(t)$ no es más difícil que una integración ordinaria cuando las dos funciones son continuas para toda t. Sin embargo, a menudo una de las funciones, o las dos, se definen en una modalidad fragmentada, y la interpretación gráfica de la convolución ilustrada en la figura 2.21 resulta especialmente útil. Las figuras 2.21*a* y *b* son las funciones consideradas aquí, por el integrado de la ecuación (1) es $v(\lambda)w(t-\lambda)$. Por supuesto, $v(\lambda)$ no es otra cosa que $v(t)$ con t reemplazada por λ, figura 2.21*c*. Pero $w(t-\lambda)$ como una función de λ se debe obtener en dos pasos: primero, $w(-\lambda)$ es $w(t)$ invertida en el tiempo con t sustituida por λ; entonces, para un valor dado de t, desplazando a $w(-\lambda)$ a la derecha t unidades se tiene $w(t-\lambda)$. La figura 2.21*d* muestra a $w(t-\lambda)$ para el caso en que $t = t_1 > 0$, ilustrando que el valor de t siempre es igual a la distancia del origen de $v(\lambda)$ al origen desplazado de $w(-\lambda)$. Por último, $v(\lambda)$ y $w(-\lambda)$ están multiplicados y el área del producto es igual a $v * w(t)$ para ese valor particular de t, figura 2.21*e*.

Conforme $v * w(t)$ se evalúa dentro del intervalo $-\infty < t < \infty$, la gráfica de $w(t-\lambda)$ se mueve de izquierda a derecha con respecto a $v(\lambda)$, y la forma real de la integración de convolución puede cambiar dependiendo del valor de t. En la figura 2.21, por ejemplo, se tiene que

$$v * w(t) = 0 \qquad t < 0$$

puesto que $(t-\lambda)$ no abarca a $v(\lambda)$ y el área del producto es cero. De manera similar

$$v * w(t) = \begin{cases} \displaystyle\int_0^t v(\lambda)w(t-\lambda)\, d\lambda & 0 < t < T \\ \displaystyle\int_{t-T}^t v(\lambda)w(t-\lambda)\, d\lambda & t > T \end{cases}$$

ya que $v(\lambda) = 0$ para $\lambda < 0$ y $w(t-\lambda) = 0$ para $\lambda < t - T$ y $\lambda > t$. Nótese que en estos casos t aparece como un límite de integración. Un simple bosquejo de las funciones consideradas nos ayudará a encontrar estos casos diferentes.

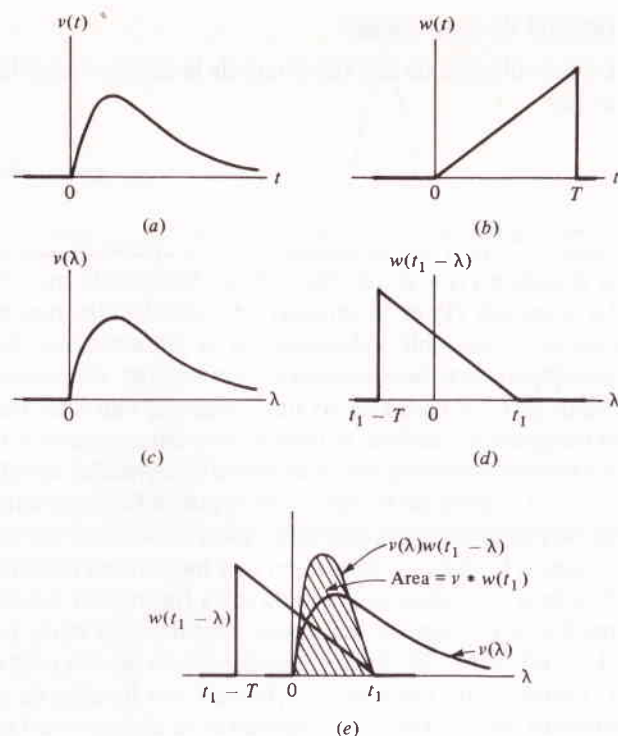



FIGURA 2.21
La interpretación gráfica de la convolución.

Un estudio más a fondo de la figura 2.21 debe indicar que $v * w(t) = w * v(t)$, es decir, obtenemos el mismo resultado invirtiendo a v y desplazándola delante de w. Esta propiedad y varias otras de convolución se listan en seguida para referencia.

$$v * w = w * v \qquad (2a)$$

$$v * (w * z) = (v * w) * z \qquad (2b)$$

$$(\alpha v + \beta w) * z = \alpha(v * z) + \beta(w * z) \qquad (2c)$$

$$\frac{d}{dt}(v * w) = v * \frac{dw}{dt} = \frac{dv}{dt} * w \qquad (3)$$

Ejemplo 2.9 Convolución de pulsos rectangulares

La convolución de dos pulsos rectangulares, figura 2.22*a*, es relativamente simple si usamos la interpretación gráfica, y el problema se divide en tres casos: $|t| > (\tau_1 + \tau_2)/2$, $(\tau_1 - \tau_2)/2 < |t| < (\tau_1 + \tau_2)/2$, y $|t| < (\tau_1 - \tau_2)/2$, suponiendo que $\tau_1 \geq \tau_2$.

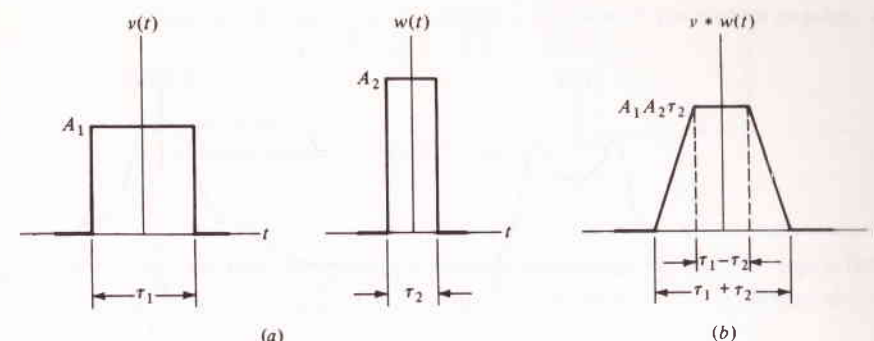



FIGURA 2.22
Convolución de pulsos rectangulares.

El resultado es una función *trapezoidal*, figura 2.22*b*, la cual degenera en una función triangular en el caso especial de que $\tau_1 = \tau_2$.

EJERCICIO 2.10 Desarrollando todos los detalles, confirme el resultado del ejemplo 2.9.

Teoremas de convolución

Habiendo definido la convolución, ahora daremos sus teoremas relativos y transformadas, los cuales son dos:

$$v * w(t) \leftrightarrow V(f)W(f) \qquad (4)$$

$$v(t)w(t) \leftrightarrow V * W(f) \qquad (5)$$

Estos teoremas establecen que la convolución en el dominio del tiempo resulta ser una multiplicación en el dominio de la frecuencia, en tanto que la multiplicación en dominio del tiempo resulta ser una convolución en el dominio de la frecuencia. Estas relaciones nos serán de utilidad por su importancia en nuestro trabajo futuro.

Para la demostración de la ecuación (4) se emplea el teorema del retardo del tiempo, como sigue:

$$\begin{aligned}\mathscr{F}[v * w(t)] &= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} v(\lambda)w(t-\lambda)\,d\lambda\right]e^{-j\omega t}\,dt \\ &= \int_{-\infty}^{\infty} v(\lambda)\left[\int_{-\infty}^{\infty} w(t-\lambda)e^{-j\omega t}\,dt\right]d\lambda \\ &= \int_{-\infty}^{\infty} v(\lambda)[W(f)e^{-j\omega\lambda}]\,d\lambda \\ &= \left[\int_{-\infty}^{\infty} v(\lambda)e^{-j\omega\lambda}\,d\lambda\right]W(f) = V(f)W(f)\end{aligned}$$

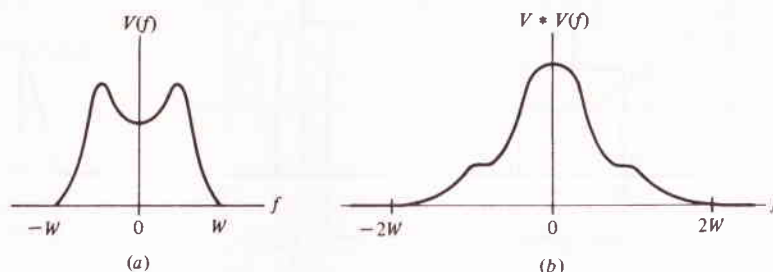



FIGURA 2.23

La ecuación (5) se puede demostrar transcribiendo la transformada de $v(t)w(t)$ y sustituyendo $w(t)$ por la integral inversa $\mathscr{F}^{-1}[W(f)]$.

EJERCICIO 2.11 Emplee la ecuación (4) con los resultados del ejemplo 2.9 para obtener la transformada de un pulso triangular, ecuación (27), sección 2.3.

Ejemplo 2.10 El espectro de $v^2(t)$

Supóngase que $v(t)$ es de banda limitada en W, con un espectro como el que se muestra en la figura 2.23*a*. ¿Cuál es entonces el espectro de $v^2(t)$? De la ecuación (5) debemos convolucionar a $V(f)$ consigo misma; el resultado será algo como lo mostrado en la figura 2.23*b*. Sin algún conocimiento específico más a fondo de $v(t)$, llegamos a esta importante conclusión: cuando $v(t)$ es de banda limitada en W, $v^2(t)$ es de banda limitada en $2W$. (Nótese la diferencia entre esta operación y el teorema de la modulación, ambos con doble ancho espectral.) El proceso puede ser iterativo para $v^3(t)$, etc., con conclusiones predecibles.

Impulso unitario

Con los ejemplos previos se ha demostrado que la convolución es una operación *suavizante*; es decir, el resultado es "más suave", "mayor" o "más amplio" que una y otra de las funciones consideradas. Pero hay una excepción notable a esta regla, o sea cuando una de las funciones es un *impulso unitario* o *función delta de Dirac* $\delta(t)$; suponiendo en este caso, que $v(t)$ es continua,

$$[v(t)] * [\delta(t)] = v(t) \qquad (6)$$

así, convolucionando con un impulso simplemente se reproduce la otra función en su totalidad.

Realmente, $\delta(t)$ no es una función en el sentido estrictamente matemático; más bien, es un miembro de esa clase especial conocida como *funciones* o *distribuciones generalizadas*.

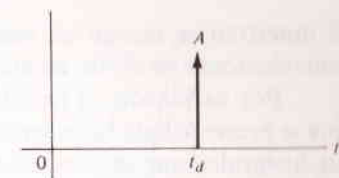



FIGURA 2.24
La representación gráfica de $A\delta(t - t_d)$.

Y como ella no es una función, el impulso unitario se define por medio de una regla de asignación o proceso en lugar de una ecuación convencional. De manera específica, dada una función ordinaria cualquiera *v(t)* que es continua en $t = 0$, $\delta(t)$ se define por

$$\int_{t_1}^{t_2} v(t)\,\delta(t)\,dt = \begin{cases} v(0) & t_1 < 0 < t_2 \\ 0 & \text{de otra manera} \end{cases} \qquad (7)$$

una regla que asigna un número, —ya sea $v(0)$ o 0— al proceso señalado en el segundo miembro.

Haciendo $v(t) = 1$ para toda t, se sigue de la ecuación (7) que

$$\int_{-\infty}^{\infty} \delta(t)\,dt = \int_{-\epsilon}^{\epsilon} \delta(t)\,dt = 1 \qquad (8a)$$

lo cual puede interpretarse diciendo que $\delta(t)$ tiene un *área unitaria* concentrada en el punto discreto $t = 0$ y no un área neta en otra parte. Desarrollando más este argumento sugiere que

$$\delta(t) = 0 \qquad t \neq 0 \qquad (8b)$$

Las ecuaciones (8*a*) y (8*b*) son las "definiciones" más familiares del impulso, y conducen a la representación gráfica común. Por ejemplo, la gráfica de $A\delta(t - t_d)$ se muestra en la figura 2.24, donde la letra A señalada por la punta de flecha significa que $A\delta(t - t_d)$ tiene área o peso A localizados en $t = t_d$. Se debe notar que la ecuación (8*b*) no es una regla de asignación o transferencia, y la teoría de la distribución, interpretada de manera estricta, no especifica otros valores para el impulso que en el sentido integral. Es, sin embargo, consistente con la ecuación (7) y nos ayuda a visualizar las propiedades del impulso bajo la operación de integración.

Dos de las propiedades más importantes de la integración son

$$\int_{-\infty}^{\infty} v(t)\,\delta(t - t_d)\,dt = v(t_d) \qquad (9)$$

$$[v(t)] * [\delta(t - t_d)] = v(t - t_d) \qquad (10)$$

ambas expresiones se pueden deducir de la ecuación (7). La ecuación (9) se conoce como la propiedad de *muestreo*, puesto que la operación indicada escoge los valores de $v(t)$ en $t = t_d$ donde "se localice" $\delta(t - t_d)$. Por otra parte, convolucionando $v(t)$ con $\delta(t - t_d)$ se tiene una propiedad *de réplica*, puesto que, de acuerdo con la ecuación (10), reproduce la función $v(t)$ completa desplazada en t_d unidades. La diferencia entre las ecuaciones (9) y (10) debe quedar perfectamente comprendida: con

el muestreo se escoge un valor particular, es decir, un número, en tanto que al convolucionar se repite completamente la función.

Por definición, el impulso no tiene significado matemático o físico, a menos que se presente bajo la integración. Aún así, es conveniente establecer tres relaciones no integrales que pueden hacerse como simplificaciones antes de la integración, puesto que son consistentes con lo que sucedería después de la integración. De manera específica, en vista de la propiedad del muestreo (9), podemos en la misma forma sustituir a $v(t)$ por $v(t_d)$ de tal manera que

$$v(t)\,\delta(t-t_d) = v(t_d)\,\delta(t-t_d) \qquad (11)$$

cuya justificación se origina de la integración de ambos miembros de $-\infty < t < \infty$. Análogamente se puede justificar la relación del cambio de escala

$$\delta(at) = \frac{1}{|a|}\,\delta(t) \qquad a \neq 0 \qquad (12)$$

de lo cual se dice que, en relación con la variable independiente t, $\delta(at)$ es un impulso que tiene un peso $1/a$. El caso especial de $a = -1$ indica la propiedad de simetría par $\delta(t) = \delta(-t)$. Por último, relacionando el impulso unitario con la función $u(t)$ de paso o escalón unitario definida en la ecuación (13), sección 2.3, vemos de la ecuación (7) que

$$\int_{-\infty}^{t} \delta(\lambda)\,d\lambda = \begin{cases} 1 & t > 0 \\ 0 & t < 0 \end{cases}$$
$$= u(t)$$

Diferenciando ambos lados se llega a

$$\delta(t) = \frac{du(t)}{dt} \qquad (13)$$

lo cual no es una definición de $\delta(t)$ sino una consecuencia de la regla de asignación (7).

Aunque un impulso no existe físicamente, existen numerosas funciones convencionales que tienen todas las propiedades de $\delta(t)$ en el límite cuando algún parámetro ϵ tiende a cero. En particular, si la función $\delta_\epsilon(t)$ es tal que

$$\lim_{\epsilon\to 0} \int_{-\infty}^{\infty} v(t)\,\delta_\epsilon(t)\,dt = v(0) \qquad (14a)$$

entonces podemos decir que

$$\lim_{\epsilon\to 0} \delta_\epsilon(t) = \delta(t) \qquad (14b)$$

Dos funciones que satisfacen la ecuación (14a) son

$$\delta_\epsilon(t) = \frac{1}{\epsilon}\,\Pi\!\left(\frac{t}{\epsilon}\right) \qquad (15)$$

$$\delta_\epsilon(t) = \frac{1}{\epsilon}\,\mathrm{senc}\,\frac{t}{\epsilon} \qquad (16)$$

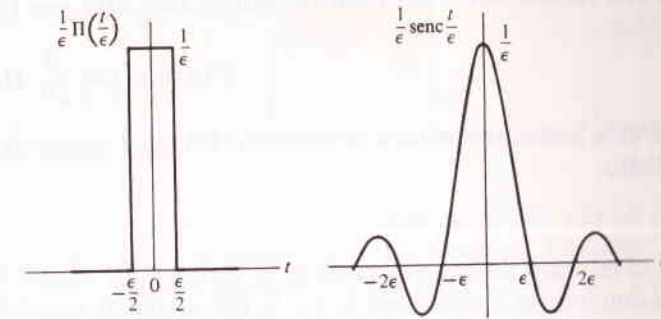



FIGURA 2.25
Dos funciones que se convierten en impulsos conforme $\epsilon \to 0$.

cuyas gráficas están en la figura 2.25. Podemos demostrar que la ecuación (15) satisface la ecuación (14a) al hacer la expansión de $v(t)$ en serie de Maclaurin antes de integrar. Se dará en breve un argumento para le ecuación (16), cuando consideremos impulsos y transformadas.

EJERCICIO 2.12 Emplee la ecuación (16) para comprobar que

$$\lim_{F\to\infty} \int_{-F}^{F} e^{\pm j2\pi ft}\,df = \delta(t) \qquad (17)$$

Impulsos en frecuencia

Las secciones 2.2 y 2.3 aclararon la diferencia entre señales de potencia periódicas y señales de energía no periódicas, siendo las de una clase descritas por medio de espectros de líneas y las de la otra por expectros continuos. Esto significa que si tenemos una señal formada de partes periódicas y no periódicas, tendremos una cierta incertidumbre, ya que se requerirían diferentes representaciones en el dominio de la frecuencia. Esta incertidumbre se resuelve si damos a los impulsos en el dominio de la frecuencia, la representación de componentes de frecuencia discreta. Tales impulsos derivan de operaciones que limitan con los pares transformados de Fourier convencionales y pueden ser designados en forma burda como *transformadas en el límite*. Nótese, sin embargo, que las correspondientes funciones del tiempo son señales de potencia que tienen energía infinita o indefinida, y el concepto de densidad espectral de energía ya no se aplica.

Como punto de partida para esta discusión, considérese la señal $v(t) = A$, una constante todo el tiempo. En relación con la figura 2.17a, sea $v(t)$ un *pulso senc* con $W \to 0$, es decir,

$$v(t) = \lim_{W\to 0} A\,\mathrm{senc}\,2Wt = A$$

Pero ya tenemos el par transformado A senc $2Wt \longleftrightarrow (A/2W)\Pi(f/2W)$, así,

$$\mathscr{F}[v(t)] = \lim_{W \to 0} \frac{A}{2W} \Pi\left(\frac{f}{2W}\right) = A\,\delta(f)$$

donde hemos recurrido a la ecuación (15) con t sustituida por f y ϵ por $2W$. Por lo tanto,

$$A \leftrightarrow A\delta(f) \qquad (18)$$

y el espectro de una constante en el dominio del tiempo viene a ser un impulso en el dominio de la frecuencia en $f = 0$. Este resultado está de acuerdo con lo dicho en el sentido de que una señal constante no tiene variación con el tiempo y su contenido espectral debe estar confinado en $f = 0$. La forma impulsiva simplemente resulta porque usamos la integración para regresar al dominio del tiempo, por medio de la transformada inversa, y se requiere de un impulso para concentrar un área diferente de cero en un punto discreto en frecuencia. Verificando este argumento en forma matemática se tiene

$$\mathscr{F}^{-1}[A\,\delta(f)] = \int_{-\infty}^{\infty} A\,\delta(f)e^{j2\pi ft}\,dt = Ae^{j2\pi ft}\Big|_{f=0} = A$$

lo cual justifica la ecuación (18) para nuestros propósitos. Nótese que el impulso ha sido integrado para obtener una cantidad física, es decir, la señal $v(t) = A$.

Como un procedimiento alterno, del anterior, y para derivar la ecuación (18), podíamos haber empezado con un pulso rectangular, $A\Pi(t/\tau)$, y con $\tau \to \infty$ obtener una constante todo el tiempo. Entonces, puesto que $\mathscr{F}[A\Pi(t/2)] = A\tau$ senc $f\tau$, y de conformidad con la ecuación (18) se requiere que

$$\lim_{\tau \to \infty} A\tau \text{ senc } f\tau = A\delta(f)$$

Esto apoya la afirmación anterior de la ecuación (16) de que una función senc viene a ser un impulso bajo condiciones de límites apropiados.

Para generalizar la ecuación (18), la aplicación directa de los teoremas de la traslación de frecuencia y la modulación nos conducen a

$$Ae^{j\omega_c t} \leftrightarrow A\delta(f - f_c) \qquad (19)$$

$$A\cos(\omega_c t + \theta) \leftrightarrow \frac{Ae^{j\theta}}{2}\delta(f - f_c) + \frac{Ae^{-j\theta}}{2}\delta(f + f_c) \qquad (20)$$

Así, el espectro de un fasor sencillo es un impulso en $f = f_c$, en tanto que el espectro de una sinusoide tiene dos impulsos, figura 2.26. Yendo aún más lejos, si $v(t)$ es una señal periódica arbitraria cuya serie de Fourier exponencial es

$$v(t) = \sum_{n=-\infty}^{\infty} c_v(nf_0)e^{j2\pi nf_0 t} \qquad (21a)$$

entonces su transformada de Fourier es

$$V(f) = \sum_{n=-\infty}^{\infty} c_v(nf_0)\,\delta(f - nf_0) \qquad (21b)$$

donde la superposición nos permite transformar la suma término a término.

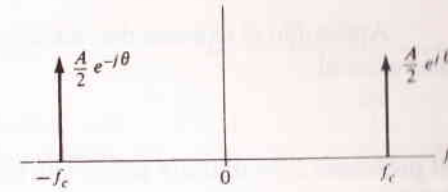



FIGURA 2.26
El espectro de $A\cos(\omega_c t + \theta)$.

Por ahora debe ser obvio que con base en las ecuaciones de la (18) a la (21), cualquier espectro de líneas de doble lado se puede convertir a espectro "continuo" empleando esta regla: convertir las líneas espectrales a impulsos cuyos pesos son iguales a las alturas de las líneas. La porción de fase del espectro de líneas se absorbe en virtud de que los pesos de los impulsos se hacen números complejos; por ejemplo, los pesos en la ecuación (21 b) son $c_v(nf_0) = c_v(nf_s)\, e^{j\arg[c_v(nf_0)]}$. Por lo tanto, con ayuda de las transformadas en el límite, podemos representar tanto a las señales periódicas como a las no periódicas por medio de espectros continuos. Además, los teoremas transformados desarrollados en la sección 2.3 pueden ahora ser aplicados a señales periódicas. Esa función sumamente extraña, la función impulso, por lo tanto,surge como la clave para unificar el análisis espectral.

Mas uno se puede preguntar: ¿Cuál es la diferencia entre el espectro de líneas y el espectro "continuo" de una señal periódica? Obviamente, no puede haber diferencia física; la diferencia cae dentro del terreno de las convenciones matemáticas. Para regresar al dominio del tiempo a partir del espectro de líneas, sumamos los fasores que las líneas representan. Para regresar al dominio del tiempo a partir del espectro continuo, integramos los impulsos para obtener fasores.

EJERCICIO 2.13 Demostrar la ecuación (20) desarrollando la transformada inversa del segundo miembro.

Impulsos en el tiempo

El impulso en el dominio del tiempo puede parecer algo poco refinado como modelo de señal, pero en la siguiente sección se verán condiciones en las que se le encuentra tanto razonable como sumamente útil. Aquí, nos concierne la transformada

$$A\delta(t) \leftrightarrow A \qquad (22)$$

la cual se deduce por transformación de Fourier al emplear la ecuación (17) para evaluar la integral. Puesto que la transformada del impulso en el tiempo tiene *amplitud constante*, su espectro contiene todas las frecuencias en igual proporción.

Se habrá observado que la ecuación (22) es la expresión dual de $A \longrightarrow A\delta(f)$. Esta relación dual tiene sus raíces en extensión recíproca, siendo las ecuaciones (18) y (22) los dos extremos; es decir, una señal constante de duración infinita tiene ancho espectral "cero", mientras que un impulso en el tiempo tiene duración "cero" y ancho espectral infinito.

Aplicando el teorema del retardo en el tiempo a la ecuación (22) se llega al par más general

$$A\delta(t - t_d) \leftrightarrow A e^{-j\omega t_d} \tag{23}$$

El problema 2.34 muestra cómo este par, con el teorema de la diferenciación, proporciona un método breve para encontrar algunas otras transformadas.

2.5 RESPUESTA DEL SISTEMA Y FILTROS

Regresemos al problema de entrada-salida o de la respuesta del sistema, como se expuso al principio del capítulo, salvo que ahora consideremos que la entrada $x(t)$ sea más o menos arbitraria. Continuaremos suponiendo que el sistema es lineal, invariante en el tiempo y asintóticamente estable, y añadiremos la limitación de que no hay energía almacenada en el sistema al aplicar la entrada. La respuesta del sistema $y(t)$ se formulará empleando tanto el análisis en el dominio del tiempo, como el análisis en el dominio de la frecuencia.

Respuesta al impulso y análisis en el dominio del tiempo

En la teoría del sistema lineal, la *respuesta al impulso* $h(t)$ de un sistema se define como la salida que resulta cuando la entrada es un impulso unitario, es decir,

$$h(t) \triangleq y(t) \quad \text{cuando } x(t) = \delta(t) \tag{1}$$

La respuesta a una entrada arbitraria $x(t)$ se encuentra entonces convolucionando $h(t)$ con $x(t)$; así

$$y(t) = h * x(t) = \int_{-\infty}^{\infty} h(\lambda)x(t - \lambda)\, d\lambda \tag{2}$$

Esta ecuación (2), a menudo conocida como la *integral de superposición*, es la base del análisis del sistema en el dominio del tiempo. Este método requiere, por lo tanto, que conozcamos la respuesta de impulso tan bien como la habilidad para efectuar la convolución.

Se tienen varias técnicas para encontrar la respuesta al impulso de un sistema, dado su modelo matemático. Si el modelo es un simple diagrama a bloques sin lazos (o circuitos) de realimentación, $h(t)$ se puede encontrar generalmente por inspección y recurriendo a la definición (1). Otras veces es más fácil calcular la *respuesta* al escalón $y_u(t)$, que se define por

$$y_u(t) \triangleq y(t) \quad \text{cuando } x(t) = u(t) \tag{3a}$$

donde

$$h(t) = \frac{dy_u(t)}{dt} \tag{3b}$$

Esto procede puesto que $y_u(t) = [h(t)] * [u(t)]$ así $dy_u(t)/dt = [h(t)] * [du(t)/dt] = [h(t)] * [\delta(t)] = h(t)$, las ecuaciones (3), (6) y (13) de la sección 2.4. Cuando estos

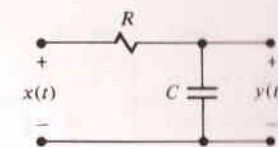



FIGURA 2.27
Filtro pasabajas RC.

métodos no funcionan, probablemente el problema es una cantidad al análisis en el dominio de la frecuencia.

Ejemplo 2.11 Respuesta de un filtro pasabajas RC, al escalón, impulso y pulso

De nuevo consideremos el filtro pasabajas RC, de la figura 2.27. Si $x(t)$ es un escalón unitario, figura 2.28*a*, la respuesta al escalón es la bien conocida

$$y_u(t) = (1 - e^{-t/RC})u(t) \tag{4a}$$

conforme a la gráfica de la figura 2.28*b*. La diferenciación nos lleva entonces a la figura 2.28*c*, o sea,

$$h(t) = \frac{dy_u(t)}{dt} = \frac{1}{RC} e^{-t/RC} u(t) \tag{4b}$$

Ahora consideremos la respuesta en un pulso rectangular de duración τ, que comienza en $t = 0$, de decir, $x(t) = a\Pi[(t - \tau/2)/\tau]$. Poniendo $x(t)$ y $h(t)$ en la ecuación (2), se tiene, con la ayuda de la interpretación gráfica de la convolución,

$$y(t) = \begin{cases} 0 & t < 0 \\ A(1 - e^{-t/RC}) & 0 < t < \tau \\ A(1 - e^{-\tau/RC})e^{-(t-\tau)/RC} & t > \tau \end{cases} \tag{5}$$

lo cual está expresado en la gráfica de la figura 2.29 para dos valores de τ/RC.

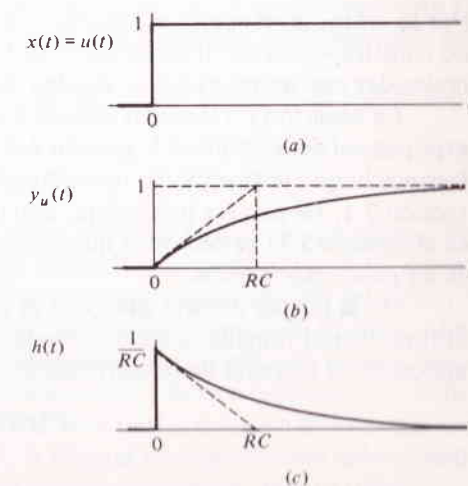



FIGURA 2.28
Formas de onda para el filtro pasabajas RC, a) entrada de escalón unitario; b) respuesta al escalón $y_u(t)$; respuesta al impulso $h(t) = dy_u(t)/dt$.

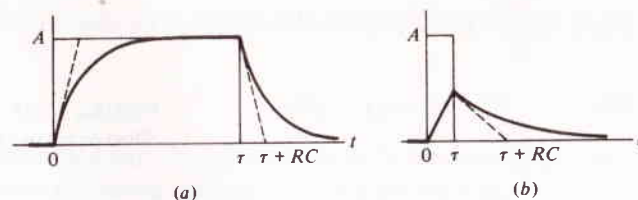



FIGURA 2.29
Respuesta a un pulso rectangular, de un filtro pasabajas RC, *a*) $\tau/\mathrm{RC} \gg 1$; *b*) $\tau/\mathrm{RC} \ll 1$.

Función de transferencia y análisis en el dominio de la frecuencia

Enlazando el análisis en el dominio del tiempo con el dominio de la frecuencia, sea $x(t) = e^{j2\pi ft}$ la entrada en la integral de superposición (2) para $-\infty < t < \infty$. Entonces,

$$y(t) = \int_{-\infty}^{\infty} h(\lambda)e^{j2\pi f(t-\lambda)}\,d\lambda$$

$$= \left[\int_{-\infty}^{\infty} h(\lambda)e^{-j2\pi f\lambda}\,d\lambda\right]e^{j2\pi ft} \tag{6}$$

donde reconocemos en la expresión entre paréntesis rectangulares, a la integral transformada de Fourier. Para identificar esa integral, recordemos que la función de transferencia $H(f)$ de un sistema ha sido definida con anterioridad, de tal manera que cuando $x(t) = e^{j2\pi ft}$, $y(t) = H(f)e^{j2\pi ft}$ —ya que la ecuación (6) tiene precisamente esta forma si tomamos

$$H(f) = \mathscr{F}[h(t)] = \int_{-\infty}^{\infty} h(t)e^{-j2\pi ft}\,dt \tag{7}$$

Por lo tanto, la respuesta al impulso y la función de transferencia de un sistema dado constituyen un par transformado de Fourier, y todas las propiedades de $H(f)$ establecidas con anterioridad se pueden deducir de la ecuación (7).

La ecuación (7) también explica por qué la transformada de Fourier del pulso exponencial del ejemplo 2.5, sección 2.3, vuelve a tener la misma dependencia de la frecuencia que la función de transferencia de un filtro pasabajas RC, ejemplo 2.1, sección 2.1. De manera imprevista, esto demuestra el par $h(t) \longrightarrow H(f)$, puesto que en el ejemplo 2.11 se demostró que la respuesta al impulso de un filtro pasabajas RC es un pulso exponencial.

De la misma manera que para el análisis del dominio de la frecuencia por sí mismo, hemos tomado la transformada de Fourier de la integral de superposición y aplicando el teorema de la convolución, ecuación (4), sección 2.4. Así,

$$\mathscr{F}[y(t)] = \mathscr{F}[h * x(t)] = \mathscr{F}[h(t)]\mathscr{F}[x(t)]$$

$$Y(f) = H(f)X(f) \tag{8}$$

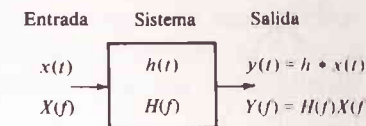



FIGURA 2.30
Relaciones de entrada y salida para un sistema lineal invariante en el tiempo.

como se ilustra esquemáticamente en la figura 2.30 con la relación al dominio del tiempo. Esta ecuación simple y elegante, base del análisis del dominio de la frecuencia, dice que el espectro de salida $Y(f)$ es igual al espectro de entrada $X(f)$ multiplicado por la función de transferencia $H(f)$. Las amplitudes y fases de los correspondientes espectros son

$$|Y(f)| = |H(f)|\,|X(f)|$$
$$\arg[Y(f)] = \arg[X(f)] + \arg[H(f)] \tag{9}$$

lo cual debe ser comparado con el resultado de estado permanente periódico, de la ecuación (23), sección 2.2. De manera adicional, si la salida *y(t)* es una señal de energía, su densidad espectral de energía y la energía total están dadas por

$$|Y(f)|^2 = |H(f)|^2\,|X(f)|^2 \tag{10}$$

$$E_y = \int_{-\infty}^{\infty} |H(f)|^2\,|X(f)|^2\,df \tag{11}$$

del teorema de la energía de Rayleigh.

Se obtiene una mejor interpretación de la ecuación (8) si tomamos a $x(t)$ como un impulso unitario; entonces, puesto que $X(f) = \mathscr{F}[\delta(t)] = 1$,

$$Y(f) = H(f) \quad \text{cuando } x(t) = \delta(t) \tag{12}$$

de conformidad con el hecho de que la función de transferencia es la transformada de la respuesta al impulso. Visto desde el dominio de la frecuencia, el espectro de la señal de entrada tiene todas las componentes de frecuencia en igual proporción en este caso, de tal manera que el espectro de salida está modulado completamente por la función de transferencia $H(f)$.

Por último, podemos regresar al dominio del tiempo tomando la transformada inversa de la ecuación (8), es decir,

$$y(t) = \mathscr{F}^{-1}[H(f)X(f)] = \int_{-\infty}^{\infty} H(f)X(f)e^{j2\pi ft}\,df \tag{13}$$

Aparece un contraste entre las ecuaciones (13) y (8), en el sentido de que el espectro de salida se obtiene de manera más fácil que la función del tiempo de salida, siendo realmente el caso si $H(f)$ y $X(f)$ son conocidas. La potencia del análisis del dominio de la frecuencia descansa en la simple relación de espectros de entrada y salida. Además, un ingeniero de comunicaciones experimentado puede a menudo deducir de su espectro, todo lo que necesita conocer. (La mayor parte de este capítulo se ha enfocado a la ejecución de tales deducciones.) Por otra parte, si se van a investigar los detalles específicos de la función del tiempo, la integral de superposición debe presentar menor problema que la ecuación (13) en lo que a encontrar $y(t)$ se refiere.

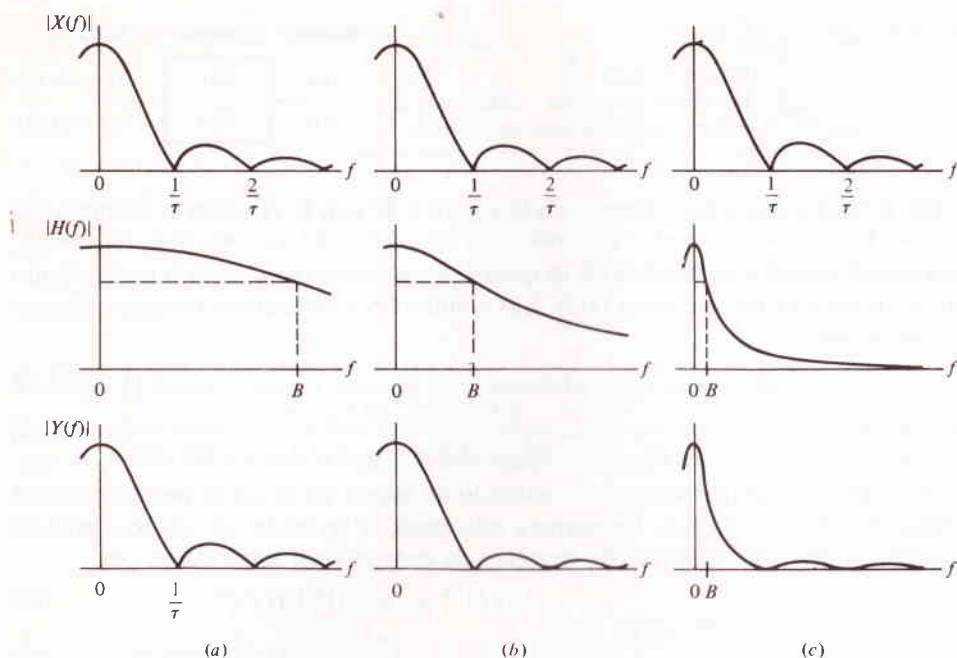



FIGURA 2.31
Análisis en el dominio de la frecuencia de la respuesta de un filtro pasabajas RC a un pulso rectangular. *a)* $B \gg 1/\tau$; *(b)* $B \approx 1/\tau$; *(c)* $B \ll 1/\tau$.

Ejemplo 2.12

Para ilustrar cuán lejos se puede ir sólo en términos del dominio de la frecuencia, supongamos un pulso rectangular de duración τ aplicado a un filtro pasabajas RC con $B = 1/2\pi RC$. Nuestros estudios previos nos han mostrado que mucho del contenido espectral del pulso está en $|f| < 1/\tau$, en tanto que el filtro responde en forma primaria a frecuencias en el intervalo $|f| < B$. Es claro que la envoltura del espectro de salida —y por lo tanto, la señal de salida— depende de los valores relativos de $1/\tau$ y de B. En la figura 2.31 se ven las gráficas de $|X(f)|$, $|H(f)|$, y $|Y(f)| = |H(f)|\,|X(f)|$ para los tres casos $B \gg 1/\tau$, $B \approx 1/\tau$ y $B \ll 1/\tau$.

En el primer caso, figura 2.31*a*, el filtro pasa todas las componentes significativas de frecuencia, puesto que $H(f) \approx 1$ para $|f| < 1/\tau$. Por lo tanto, $Y(f) \approx X(f)$ y $y(t) = X(t)$, de tal manera que la función del tiempo de salida debe verse muy semejante a la de entrada. Para una comprobación directa de esta conclusión, la figura 2.29*a* muestra las formas de onda reales bajo esta condición. En el segundo caso, figura 2.31*b*, la forma del espectro de salida depende tanto de $X(f)$ como de $H(f)$, de manera que $y(t)$ diferirá sustancialmente de $x(t)$. Decimos entonces que la salida está *distorsionada* en el sentido de que no se asemeja a la entrada, pero una exposición más precisa ocasiona cálculos en el dominio del tiempo. En el tercer caso, figura 2.31*c*, el espectro de entrada es constante o "plano" sobre $|f| < B$ de tal manera

que $Y(f) \approx A\tau H(f)$. Puesto que el espectro de salida ahora depende en forma primaria de $H(f)$, la señal de salida se verá semejante a la *respuesta al impulso* del filtro, es decir, $y(t) \approx A\tau h(t)$. De nuevo, esto se confirma con el resultado en el dominio del tiempo en la gráfica de la figura 2.29*b*.

Extrapolando este último caso a otros filtros y formas de pulsos, establecemos la siguiente regla: si el espectro de salida es esencialmente constante sobre la banda de frecuencia, donde tiene respuesta significativa, entonces la señal de salida es en esencia la respuesta al impulso del filtro. Bajo tales condiciones, es muy razonable tomar como modelo un impulso como señal de entrada.

Conexiones en paralelo, en cascada y realimentadas

En la mayoría de las ocaciones, un sistema de comunicación comprende muchas unidades o subsistemas interconectados. Cuando los subsistemas en cuestión se describen por medio de funciones de transferencia individuales, es posible y deseable agruparlos y hablar de la función de transferencia en su conjunto. Las relaciones correspondientes a dos subsistemas concertados en paralelos, en cascada y realimentados, se dan en seguida. Las configuraciones más complicadas se pueden analizar por la aplicación sucesiva de estas reglas básicas. Sin embargo, debe hacerse una suposición esencial, y es que se ha estimado alguna interacción o efectos de *carga* de las funciones de transferencia individuales, de tal manera que éstas representan la respuesta real del subsistema en el contexto del sistema completo.

En la figura 2.32*a* muestra dos subsistemas en *paralelo*; ambos tienen la misma entrada y sus salidas se suman para obtener la salida del sistema. De la superposición se sigue que $Y(f) = [H_1(f) + H_2(f)]X(f)$ de tal manera que la función de transferencia total es

$$H(f) = H_1(f) + H_2(f) \quad \text{Conexión en paralelo} \quad (14)$$

En la conexión en *cascada*, figura 2.32*b*, la salida de la primera unidad es la entrada de la segunda; así, $Y(f) = H_2(f)[H_1(f)X(f)]$ y

$$H(f) = H_1(f)H_2(f) \quad \text{Conexión en cascada} \quad (15)$$

La conexión realimentada, figura 2.32*c*, difiere de las otras dos en que la salida se regresa a través de $H_2(f)$ y es sustraída de la entrada. Así, $Y(f) = H_1(f)[X(f) - H_2(f)Y(f)]$ y reacomodando se tiene $Y(f) = \{H_1(f)/[1 + H_1(f)H_2(f)]\}X(f)$ de tal manera que

$$H(f) = \frac{H_1(f)}{1 + H_1(f)H_2(f)} \quad \text{Conexión realimentada} \quad (16)$$

A este caso se le designa en forma más apropiada como conexión de realimentación *negativa* para diferenciarlo de la realimentación positiva, donde la señal regresada se agrega a la entrada en vez restarla.

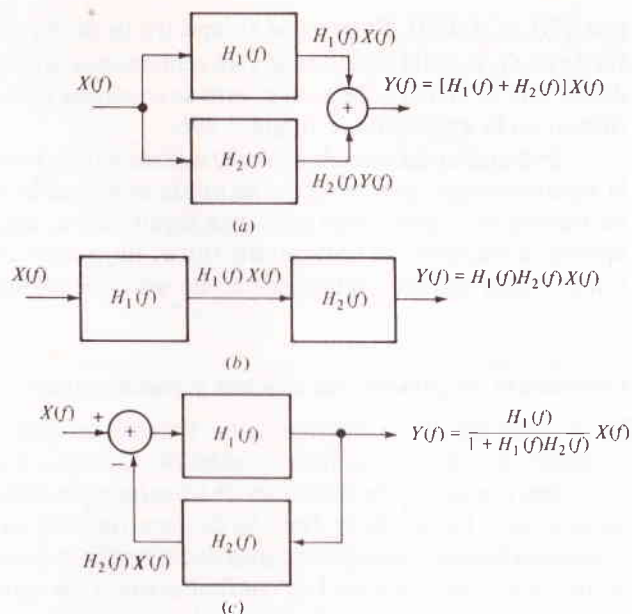



FIGURA 2.32
a) Conexión en paralelo; *b*) conexión en cascada; *c*) conexión realimentada.

Ejemplo 2.13 Retención de orden cero

El sistema de retención de orden cero, figura 2.33*a*, tiene varias aplicaciones en la comunicación eléctrica. Aquí se toma como un ejercicio instructivo de las relaciones paralelo y cascada. Pero primero necesitamos las funciones de transferencia individuales, las que se determinan como sigue: la rama superior de la sección paralelo es una trayectoria siempre recta, de tal manera que, trivialmente, $H_1(f) = 1$; la rama inferior genera un retardo puro en el tiempo de T segundos e inversión de signo y agrupándolos se tiene $H_2(f) = e^{-j2\pi ft}$ por aplicación del teorema del retardo en el tiempo; usando el teorema de la integración, el integrador en el bloque final tiene $H_3(f) = 1/j2\pi f$. La figura 2.33*b* es el diagrama a bloques equivalente en términos de estas funciones de transferencia.

En adelante el resto del trabajo és fácil. Combinamos las ramas en paralelo en $H_{12}(f) = H_1(f) + H_2(f)$ y usando la regla para cascada obtenemos

$$\begin{aligned} H(f) &= H_{12}(f)H_3(f) = [H_1(f) + H_2(f)]H_3(f) \\ &= [1 - e^{-j2\pi fT}]\frac{1}{j2\pi f} \\ &= \frac{e^{j\pi fT} - e^{-j\pi fT}}{j2\pi f} e^{-j\pi fT} = \frac{\operatorname{sen} \pi fT}{\pi f} e^{-j\pi fT} \\ &= T \operatorname{senc} fT e^{-j\pi fT} \end{aligned} \tag{17}$$

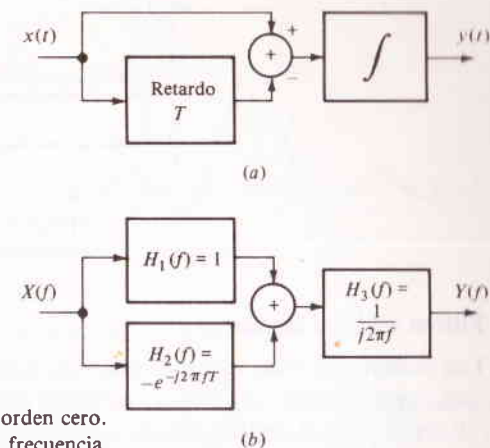



FIGURA 2.33
Diagrama a cuadros de una retención de orden cero.
a) Dominio del tiempo; *b*) dominio de la frecuencia.

Por lo que hemos llegado al resultado no usual de que la relación de amplitudes de este sistema, ¡es una *función senc* en frecuencia!

Para confirmar este resultado en otra forma, calculemos la respuesta al impulso $h(t)$ recurriendo a la definición de que $y(t) = h(t)$ cuando $x(t) = \delta(t)$. La inspección de la figura 2.33*a* nos hace ver que la entrada en el integrador es entonces $x(t) - x(t - T) = \delta(t) - \delta(t - T)$, y así

$$\begin{aligned} h(t) &= \int_{-\infty}^{t} [\delta(\lambda) - \delta(\lambda - T)]\, d\lambda = \begin{cases} 1 & 0 < t < T \\ 0 & \text{de otra manera} \end{cases} \\ &= \Pi\left(\frac{t - T/2}{T}\right) \end{aligned} \tag{18}$$

y la respuesta al impulso es un pulso rectangular, figura 2.34. El lector puede tener alguna dificultad para identificar la ecuación (17) como la transformada de la ecuación (18).

EJERCICIO 2.14 Deduzca la ecuación (17) haciendo $x(t) = e^{j2\pi ft}$ en la figura 2.33*a* y encuentre $y(t)$.

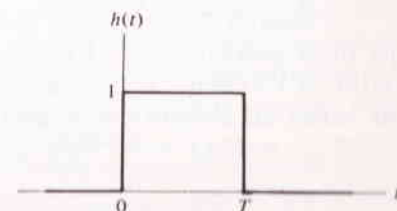



FIGURA 2.34
Respuesta al impulso de una retención de orden cero.

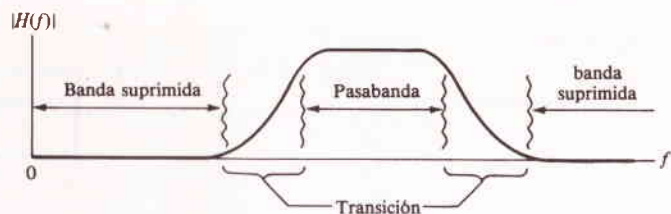



FIGURA 2.35
Pasabanda y banda suprimida de un filtro pasabanda típico.

Filtros reales e ideales

Los sistemas o redes que presentan características selectivas de frecuencia, se designan como *filtros*. Un filtro pasabajas (LPF), por ejemplo, pasa sólo frecuencias "bajas" —en el sentido de que su relación de amplitudes $|H(f)|$ es mucho mayor a frecuencias bajas que a frecuencias altas. El filtro pasabajas RC y la retención de orden cero recién discutida, ilustran este tipo de respuesta en frecuencia. De manera similar, existen filtros *pasaaltas* (HPF), filtros *pasabanda* (BPF) y filtros *supresores de banda*. El circuito sintonizado del multiplicador de frecuencia actúa como un BPF.

De manera ideal, un filtro debe tener un límite agudo entre sus bandas *pasabanda* y de *rechazo* o de supresión. Entonces el ancho de banda del filtro se podría medir de manera no ambigua como el ancho del pasabanda. Sin embargo, las curvas de la relación de amplitudes no tienen puntos de demarcación pronunciados, así, los extremos del pasabanda y de la banda de rechazo son ambiguos, por lo que existe una banda de *transición* entre ambos. La figura 2.35 ilustra esta característica en un filtro pasabanda típico; nótese que la porción de frecuencia negativa ha sido omitida y las diferentes bandas se indican sólo en términos de frecuencias positivas, dada la simetría par de $|H(f)|$ más el hecho de que las frecuencias negativas son una ficción que se introdujo por conveniencia analítica.

Al tratar con filtros reales, el ancho de banda generalmente se toma del intervalo de *frecuencias positivas* en el cual $|H(f)|$ no decrece a valores menores de $1/\sqrt{2}$ veces el valor máximo de $|H(f)|$ en el pasabanda. Esta convención particular de ancho de banda y hay otras —se conoce como la *potencia media* o *ancho de banda a 3 decibeles (dB)*. El nombre procede del hecho de que una entrada senoidal a la frecuencia de corte a 3 dB saldría con su potencia promedio reducida en $(1/\sqrt{2})^2 = \frac{1}{2}$ comparada con una sinusoide en el centro de la pasabanda. Convirtiendo esta relación a decibeles † se tiene $10 \log_{10} \frac{1}{2} \approx -3$ dB, de tal manera que la relación de potencia en el borde del pasabanda está 3 dB abajo del centro.

Como un ejemplo de esta convención, el ancho de banda de la potencia media del filtro pasabajas RC LPF (figura 2.6) es $B = \frac{1}{2}\pi RC$ puesto que $H(f)_{\text{máx}} = H(0) = 1$ y $|H(B)| = 1\sqrt{2}$; la banda de paso es $0 \leq f \leq B$. Por lo mismo, el ancho de banda del circuito sintonizado BPF en la figura 2.12 es $B = f_r/Q$ puesto que $H(f)|_{\text{máx}} = |H(f_r)| = R|\ H(f_r) \pm f_r/2Q\ | = R/\sqrt{2}$. (En este punto el lector debe

† Véase la Tabla E.

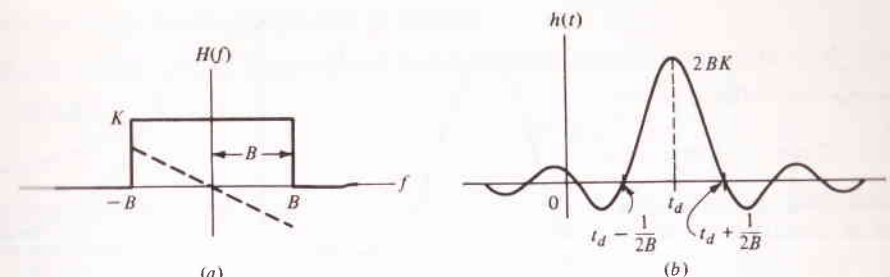



FIGURA 2.36
Filtro pasabajas ideal. *a*) función de transferencia; *b*) respuesta al impulso.

echar otro vistazo a las figuras 2.6 y 2.12 a la luz de estos comentarios, y fijarse de manera especial que B se mide en términos del intervalo de frecuencias positivas.)

Los diseños de filtros más complicados tienen características mucho más selectivas que nuestros ejemplos simples, al grado tal de que se aproximan en forma progresiva a las transiciones de frecuencia observadas en un pulso rectangular. En el límite definimos a un filtro pasabajas *ideal* como aquel que tiene una característica rectangular

$$H(f) = Ke^{-j\omega t_d}\Pi\left(\frac{f}{2B}\right) \qquad (19)$$

donde K es la amplificación, t_d el retardo en el tiempo y B el ancho de banda. En la figura 2.36*a* se trazó esta función de transferencia. Nótese que el ancho de banda es no ambiguo y que la respuesta fuera del pasabanda es idénticamente cero. Los filtros pasabajas y pasaaltas ideales quedan definidos de manera similar y los BPF aparecerán en una discusión posterior en el capítulo 5.

En teoría avanzada de redes se demuestra que los filtros ideales no se pueden realizar en forma física. Omitimos aquí la demostración general y damos en su lugar un argumento basado en la respuesta al impulso. Considérese, por ejemplo, la respuesta al impulso de un filtro pasabajas ideal. Por definición y empleando la ecuación (19),

$$h(t) = \mathscr{F}^{-1}[H(f)] = \mathscr{F}^{-1}\left[Ke^{-j\omega t_d}\Pi\left(\frac{f}{2B}\right)\right]$$
$$= 2BK \operatorname{senc} 2B(t - t_d) \qquad (20)$$

lo cual se expresó en la figura 2.36*b* puesto que $h(t)$ es la respuesta a $\delta(t)$ y $h(t)$ tiene valores diferentes de cero para, *la salida aparece antes de que la señal de entrada se aplique*. De tal filtro decimos *que se anticipa*, y —la porción de la salida que aparece antes que la entrada se le conoce como *precursora*. Sin duda, tal comportamiento es físicamente imposible y, por lo tanto, el filtro no es realizable. Se encuentran resultados similares en los casos de filtros pasabanda y pasaalta.

Los filtros ideales, aunque ficticios son, en forma conceptual, útiles en el estudio de sistemas de comunicación, y pueden diseñarse filtros prácticos que se acer-

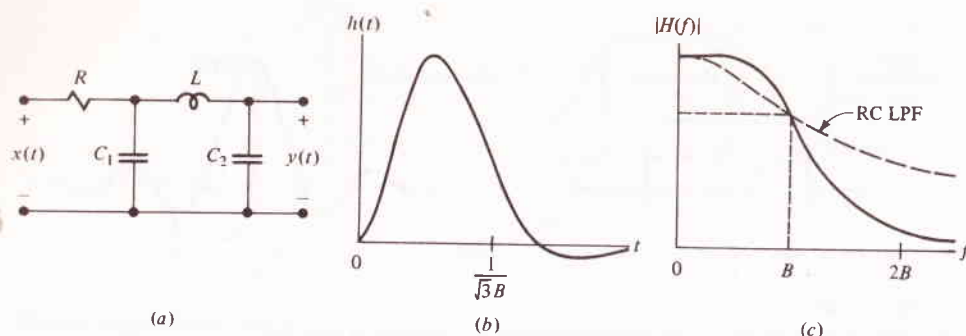



FIGURA 2.37
El filtro Butterworth de tercer orden, *a*) Circuito ($R = 100$, $C_1 = 1/400\pi B$, $C_2 = 3C_1$, $L = 200/3\pi B$); *b*) respuesta al impulso; *c*) relación de amplitudes.

quen bastante al idóneo, para propósitos de ingeniería. De hecho, como el número de elementos reactivos se incrementa sin límite, la función de transferencia se puede hacer en forma arbitraria lo más cercana a la de un filtro ideal. Pero de manera simultánea, el retardo en tiempo del filtro aumenta sin límite. Como algo colateral, observemos que el retardo infinito en el tiempo significa que las precursoras aparecerán siempre después de que se aplique la entrada, lo cual debe ser cierto en un filtro real.

Ejemplo 2.14 Filtro pasabajas Butterworth

La figura 2.37*a* es el diagrama de circuito de un filtro Butterworth de tercer orden con ancho de banda B a 3 dB. Su función de transferencia y su respuesta al impulso son

$$H(f) = \left\{1 - 2\left(\frac{f}{B}\right)^2 + j\left[2\left(\frac{f}{B}\right) - \left(\frac{f}{B}\right)^3\right]\right\}^{-1}$$

$$h(t) = 2\pi B\left[e^{-2\pi Bt} - \frac{2}{\sqrt{3}}\,e^{-\pi Bt}\cos\left(\pi\sqrt{3}\,Bt + 30°\right)\right]u(t)$$

como se muestra en la gráfica de la figura 2.37*b* y *c*. La similitud con un filtro pasabajas ideal es aparente. La figura 2.37*c* muestra también que este filtro tiene una región de transición más angosta que el filtro pasabajas RC simple.

Para el caso de M elementos reactivos, la relación de amplitudes de un Butterworth de orden M-ésimo es

$$|H(f)| = \left[1 + \left(\frac{f}{B}\right)^{2M}\right]^{-1/2} \qquad (21)$$

del cual se dice que es de aplanamiento máximo, puesto que las primeras M derivadas de $|H(f)|$ son cero en $f = 0$. (En forma incidental, el filtro pasabajas RC es un Butterworth de primer orden.) Se dan otros tipos de diseños de filtros en literatura.

Limitación de banda y limitación en tiempo

Recién dijimos que una señal $v(t)$ es ***de banda limitada***, si existe una constante W tal que

$$\mathscr{F}[v(t)] = 0 \qquad |f| > W \qquad (22)$$

es decir, el espectro no tiene contenido fuera de $|f| < W$. De manera similar, una señal *limitada en el tiempo* tiene la propiedad de que, para las constantes $t_1 < t_2$,

$$v(t) = 0 \qquad t < t_1 \quad \text{y} \quad t > t_2 \qquad (23)$$

de tal manera que la señal "empieza" en el tiempo t_1 y "termina" en el tiempo t_2. Aquí consideraremos el significado de estas dos definiciones.

Los filtros ideales y las señales de banda limitada, son conceptos que van de la mano. En realidad, una señal arbitraria que pasa a través de un filtro pasabajas ideal produce una señal de banda limitada en la salida. Pero se ha visto que la respuesta al impulso de un filtro pasabajas ideal, es un pulso senc que existe para todo el tiempo. Ahora aseguramos que cualquier señal que salga de un filtro pasabajas ideal existirá para todo el tiempo; establecido de otra manera, una señal de banda limitada no puede ser estrictamente limitada en el tiempo. A la inversa, una señal estrictamente limitada en el tiempo no puede ser de banda limitada. En pocas palabras, *la limitación de banda y la limitación en el tiempo son mutuamente incompatibles*. Una demostración general de esta afirmación es difícil y no la intentaremos aquí.† Sin embargo, cada par transformado encontrado en este capítulo es al mismo tiempo consistente; para ejemplos, véanse las figura 2.13, 2.14, 2.17, 2.20.

Estas observaciones tienen implicaciones con los modelos de la señal y del sistema utilizados en el estudio de sistemas de comunicaciones. Puesto que una señal no puede ser en forma simultánea limitada en banda y en tiempo, debemos desentendernos de las señales de banda limitada (y filtros ideales) o aceptar modelos de señales que existan todo el tiempo. Pero una señal real desde el punto de vista físico *es* estrictamente limitadas en el tiempo; ella empieza y termina, o sea, es conectada y desconectada. Por otra parte, el concepto de espectros de banda limitada es muy poderoso, y en cuanto a propósitos de ingeniería se refiere se descarta por completo.

Por fortuna, la solución del dilema en realidad no es tan difícil, para lo cual, sin embargo, hay un cierto arreglo. Aunque una señal estrictamente limitada en tiempo no es estrictamente de banda limitada, su espectro puede ser *esencialmente* cero fuera de un cierto intervalo de frecuencias, en el sentido de que las componentes de frecuencia no tomadas en cuenta contienen una porción inconsecuente de la energía total; por ejemplo, considérese $f \gg 1/\tau$ en el espectro de un pulso rectangular. De manera similar, una señal de banda estrictamente limitada puede ser virtualmente cero fuera de un cierto intervalo de tiempo; por ejemplo, senc $2Wt \approx 0$ para $t \gg 1/2W$. Por lo tanto, no es apropiado hablar de señales que son o de banda limitada o limitadas en tiempo para muchos propósitos prácticos.

† Véase Wozencraft y Jacobs (1965, apénd. 5B). (1965, Apén. 5B).

2.6 CORRELACIÓN Y DENSIDAD ESPECTRAL

En las secciones 2.3 y 2.4, se observaron dos casos particulares. Las señales de energía no periódicas se representaron en el dominio de la frecuencia por medio de transformadas de Fourier que son funciones continuas de la frecuencia, libres de impulsos; la energía de la señal es

$$E = \int_{-\infty}^{\infty} |v(t)|^2\, dt = \int_{-\infty}^{\infty} v(t)v^*(t)\, dt \tag{1}$$

y la *densidad espectral de energía* $|V(f)|^2$ da su distribución en frecuencia; la potencia promediada en todo el tiempo es cero, puesto que E es finita. Por otra parte, las señales de potencia periódicas se representaron en el dominio de la frecuencia por medio de espectros impulsivos provinientes de transformadas en el límite; la potencia promedio es

$$P = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} |v(t)|^2\, dt = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} v(t)v^*(t)\, dt \tag{2}$$

y el concepto de densidad espectral de energía no se aplica, ya que la energía total debe ser infinita cuando $P \neq 0$. Sin embargo, debe entenderse perfectamente que al hablar de distribución de la potencia en el dominio de la frecuencia, se considera descrita por una *densidad espectral de potencia.*

El objeto de esta sección es desarrollar en forma más completa el concepto de densidad espectral a medida que se aplica a los casos citados o al caso de *señales aleatorias,* —en anticipación a las necesidades del capítulo 3—. Esta generalización se obtiene al precio de emplear el punto de vista algo abstracto de *espacio de la señal*, pero el precio es una inversión de gran alcance que vale la pena, ya que nos reditúa beneficio ahora y en nuestro trabajo futuro. A fin de conseguir resultados útiles lo antes posible, muchas de las deducciones del espacio de la señal se han anotado aparte en el Apéndice A, el cual puede ser consultado para aclarar los detalles que se hayan omitido.

Producto escalar, norma y ortogonalidad

Sean $v(t)$ y $w(t)$ dos señales de la misma *clase*, es decir, cualquiera del tipo de energía o del tipo de potencia. Su *producto escalar* es una cantidad —posiblemente compleja— que se denota por $< v(t), w(t) >$ y definida por

$$\langle v(t),w(t)\rangle \triangleq \begin{cases} \displaystyle\int_{-\infty}^{\infty} v(t)w^*(t)\, dt & \text{Señales de energía} \quad (3a) \\ \displaystyle\lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2} v(t)w^*(t)\, dt & \text{Señales de potencia} \quad (3b) \end{cases}$$

† Esta notación es diferente de $\overline{v(t)}$, el *promedio en tiempo* de $v(t)$; el símbolo del producto escalar siempre tiene *dos* funciones separadas por una coma.

Si ocurre que $v(t)$ y $w(t)$ son periódicos con un periodo T_0, la ecuación (3*b*) se simplifica a

$$\langle v(t),w(t)\rangle = \frac{1}{T_0}\int_{T_0} v(t)w^*(t)\, dt \qquad \text{Señales periódicas} \tag{3c}$$

En el capítulo 3 se dará otra difinición de producto escalar, inclusive en el caso de señales aleatorias. La importancia del concepto de producto escalar es que $< v(t), w(t) >$ tiene ciertas propiedades invariantes sin importar la definición específica.

Antes de establecer algunas de estas propiedades, nótese que el producto escalar de $v(t)$ consigo mismo es igual a la energía E o a la potencia P. Generalizando, la *norma* de $v(t)$ se define por

$$\|v\| \triangleq \langle v(t),v(t)\rangle^{1/2} \tag{4}$$

una cantidad real no negativa, que no se puede confundir con la función $v(t)$. Por lo que

$$\|v\|^2 = \langle v(t), v(t)\rangle = \begin{cases} E & \text{Señales de energía} \\ P & \text{Señales de potencia} \end{cases} \tag{5}$$

como se sigue de las ecuaciones (1) a la (4).

La desigualdad de Schwarz enlaza las ecuaciones (3) y (4) en la forma

$$|\langle v(t),w(t)\rangle| \leq \|v\|\,\|w\| \tag{6a}$$

la cual establece un límite superior en la magnitud del producto escalar. El límite superior se obtiene sólo cuando las dos señales son directamente proporcionales, es decir,

$$|\langle v(t),w(t)\rangle| = \|v\|\,\|w\| \qquad \text{si } w(t) = \alpha v(t) \tag{6b}$$

donde α es una constante arbitraria. Por otra parte, supóngase que $v(t)$ y $w(t)$ son *ortogonales*, cuyo significado es

$$\langle v(t),w(t)\rangle = 0 \tag{7}$$

Bajo esta condición

$$\|v + w\|^2 = \|v\|^2 + \|w\|^2 \tag{8}$$

la demostración de esto se da en el Apéndice A con la demostración de la ecuación (6).

En la teoría del espacio de la señal, se ven las dos señales ortogonales como vectores perpendiculares y la ecuación (8) resulta equivalente al teorema de Pitágoras. En este punto interpretamos a $< v(t), w(t) > = 0$ como la condición para la *superposición de la energía* ~~*o de la potencia*~~. Por ejemplo, supongamos que $v(t)$ y

$w(t)$ son señales ortogonales de energía con $E_v = \|v\|^2$ y $E_w = \|w\|^2$; entonces la señal $z(t) = v(t) + w(t)$ tiene energía $E_z = \|v + w\|^2$ la cual, por la ecuación (8) es igual a $E_v + E_w$.

También es útil enumerar algunas relaciones específicas entre las funciones o formas de onda $v(t)$ y $w(t)$ que resultan de hecho ser ortogonales. Hay tres condiciones bajo las cuales $< v(t), w(t) > = 0$, y son:

1 Cuando $v(t)$ y $w(t)$ tienen *simetría opuesta*, es decir, una es función par y la otra es impar.
2 Cuando $v(t)$ y $w(t)$ no se traslapan o que son *desunidas en el tiempo*, es decir, una es cero cuando la otra es diferente de cero y viceversa.
3 Cuando $v(t)$ y $w(t)$ son *desunidas en frecuencia*, es decir, sus espectros no se traslapan.

Las primeras dos condiciones se deducen directamente de la ecuación (3); la ecuación (13), sección 2.3, nos lleva a la tercera. Debemos llevar en mente que éstas no son las únicas relaciones que dan lugar a la ortogonalidad.

Como detalle final, nótese que el producto escalar es la medida del grado de *similitud* entre dos señales. Si son proporcionales (es decir, similares), entonces $< v(t), w(t) >$ es máxima; si son ortogonales (es decir, desemejantes), entonces $< v(t), w(t) > = 0$.

Funciones de correlación

Para cualquiera de dos señales del mismo tipo, la *correlación cruzada* de *v(t)* con *w(t)* se define† como ecuación (a

$$R_{vw}(\tau) \triangleq \langle v(t), w(t-\tau)\rangle \tag{9}$$

un producto escalar, en el cual la segunda señal experimenta un desplazamiento o corrimiento en un tiempo τ segundos. El desplazamiento τ es arbitrario y, de hecho, la correlación tiene a τ como su variable independiente, habiendo desaparecido la variable temporal t, por razón de la operación producto escalar. Por extensión de los argumentos recien vistos, $R_{vw}(\tau)$ da la medida de la simitud entre $v(t)$ y $w(t - \tau)$ como una función del desplazamiento τ de $w(t)$ con respecto a $v(t)$. Por lo tanto, la correlación cruzada es una medida más comprensible que el producto escalar regular, porque detecta cualesquier de las similitudes desplazadas en el tiempo que deben despreciarse por $< v(t), w(t) >$.

Ahora supóngase que formamos la correlación de *v(t)* consigo misma, es decir,

$$R_v(\tau) \triangleq R_{vv}(\tau) = \langle v(t), v(t-\tau)\rangle \tag{10}$$

† También se usa la definición $v(t), w(t+\tau)$.

la cual se conoce como la *función de autocorrelación*. Físicamente, la autocorrelación tiene la misma interpretación que la correlación cruzada, salvo que compara una señal consigo misma desplazada en el tiempo. Pero esto significa que $R_v(\tau)$ nos dirá algo acerca de la *variación en el tiempo* de $v(t)$, al menos en un sentido integrado o promediado. Por ejemplo, si $|R_v(\tau)|$ es grande, entonces inferimos que $v(t - \tau)$ es muy similar (proporcional) a $\pm\, v(t)$ para ese valor particular de τ; en forma inversa, si $R_v(\tau) = 0$ para algún valor de τ, entonces sabemos que $v(t)$ y $v(t - \tau)$ son ortogonales. En breve, recurriremos a esta interpretación y a la función de densidad espectral. Necesitamos primero, no obstante, mencionar dos propiedades matemáticas de autocorrelación.

Si aplicamos las ecuaciones (5) y (6) a la ecuación (10) se tiene

$$R_v(0) = \langle v(t), v(t)\rangle = \|v\|^2 \tag{11a}$$

$$|R_v(\tau)| \leq R_v(0) \tag{11b}$$

de tal manera que $R_v(\tau)$ tiene un máximo en $\tau = 0$, donde ella es igual a la energía o potencia de la señal $\|v\|^2$. A partir de las ecuaciones (3) y (10), se puede demostrar también que

$$R_v(-\tau) = R_v^*(\tau) \tag{12}$$

Así, si $v(t)$ es *real*, $r_v(\tau)$ es *real* y con *simetría par*.

Considérese ahora la suma de dos señales, digamos $z(t) = v(t) + w(t)$, cuya correlación es $R_z(\tau) = R_v(\tau) + R_{vw}(\tau) + R_{wv}(\tau) + r_w(\tau)$. Si las señales componentes son *ortogonales para toda* τ, es decir, si

$$R_{vw}(\tau) = R_{wv}(\tau) = 0 \tag{13a}$$

entonces

$$R_z(\tau) = R_v(\tau) + R_w(\tau) \tag{13b}$$

y haciendo $\tau = 0$,

$$\|z\|^2 = \|v\|^2 + \|w\|^2 \tag{13c}$$

En este caso, se dice que las señales son *incoherentes*, y tenemos superposición de funciones de correlación, así como superposición de energía o potencia.

Ejemplo 2.15 Autocorrelación de una sinusoide

Calculemos la autocorrelación de la señal periódica de potencia

$$z(t) = A\cos(\omega_0 t + \theta) \qquad \omega_0 = 2\pi/T_0$$

Mientras que esto se puede hacer empleando en forma directa las ecuaciones (3c) y (10), es más fácil e instructivo escribir

$$z(t) = \underbrace{\frac{A}{2}e^{j\theta}e^{j\omega_0 t}}_{v(t)} + \underbrace{\frac{A}{2}e^{-j\theta}e^{-j\omega_0 t}}_{w(t)}$$

Esto resulta ser un método mejor para el caso que nos ocupa porque $v(t)$ y $w(t - \tau)$ son ortogonales para toda τ —puesto que son desunidas en frecuencia— y, más aún, $w(t) = v^*(t)$. Así, $R_{vw}(\tau) = R_{wv}(\tau) = 0$ y $R_w(\tau) = R_v^*(\tau)$, por lo que la ecuación (13*b*) viene a ser $R_z(\tau) = R_v(\tau) + R_v^* = 2 \text{ Re } [R_v(\tau)]$. Incidentalmente, aunque no todos los problemas se pueden simplificar tanto, el estudiante debe estar pendiente de tales posibilidades.

Continuando con el cálculo, tenemos

$$R_v(\tau) = \frac{1}{T_0}\int_{T_0}\left[\frac{A}{2}e^{j\theta}e^{j\omega_0 t}\right]\left[\frac{A}{2}e^{-j\theta}e^{-j\omega_0(t-\tau)}\right]dt = \frac{A^2}{4}e^{j\omega_0\tau} \tag{14}$$

y por ende

$$R_z(\tau) = 2 \text{ Re } [R_v(\tau)] = \frac{A^2}{2}\cos\omega_0\tau \tag{15}$$

de tal manera que $R_z(0) = A^2/2 = \|z\|^2$ como se predijo. Nótese que la correlación de una sinusoide es otra sinusoide de la misma frecuencia, pero más en el "dominio de τ" que en el dominio del tiempo. El parámetro fase θ desaparece obedeciendo al efecto promediante de la correlación. Por razón de este efecto, concluimos que la autocorrelación no define en forma única a la señal, por ejemplo, todas las $z(t) = A \cos(\omega_0 t + \theta)$ tienen la misma $R_z(\tau)$ prescindiendo de θ.

EJERCICIO 2.15 La autocorrelación de una señal de energía es un tipo de *convolución*, puesto que al sustituir a t por λ en la ecuación (3*a*) se tiene

$$R_v(\tau) = \int_{-\infty}^{\infty} v(\lambda)v^*(\lambda - \tau)\,d\lambda = [v(\tau)] * [v^*(-\tau)] \tag{16}$$

Emplee esto para demostrar que un pulso rectangular

$$v(t) = A\Pi\left(\frac{t}{T}\right) \tag{17a}$$

tiene una autocorrelación triangular

$$R_v(\tau) = A^2T\Lambda\left(\frac{\tau}{T}\right) \tag{17b}$$

Subrayando la condición de no unicidad observada con anterioridad, la ecuación (17*b*) también vale cuando $v(t) = A\Pi[(t - t_d)/T]$ para cualquier valor de t_d.

Funciones de densidad espectral

En vista de nuestra observación respecto de que $R_v(\tau)$ proporciona información acerca del comportamiento de $v(t)$ en el dominio del tiempo, parece conveniente investigar la transformada de Fourier de $R_v(\tau)$ como una posible base para el análisis del dominio de la frecuencia. Considérese, por lo tanto,

$$G_v(f) \triangleq \mathscr{F}[R_v(\tau)] = \int_{-\infty}^{\infty} R_v(\tau)e^{-j2\pi f\tau}\,d\tau \tag{18a}$$

a la cual se le conoce como la *función de densidad espectral*, por razones recié explicadas. Habiéndose establecido esta definición, se sigue que $\mathscr{F}^{-1}[G_v(f)]$ es igua a $R_v(\tau)$, es decir,

$$R_v(\tau) = \int_{-\infty}^{\infty} G_v(f)e^{+j2\pi f\tau}\,df \qquad (18b$$

Así, tenemos el par transformado de Fourier

$$R_v(\tau) \leftrightarrow G_v(f) \qquad (19$$

donde τ ocupa el lugar de t. La ecuación (19) recibe el nombre de *teorema d Wiener-Kinchine*. En el caso especial e importante donde $v(t)$ es *real*, $G_v(f)$ es *real par*, puesto que $R_v(\tau)$ es real y par.

La propiedad fundamental de $G_v(f)$ es que integrándola sobre toda la frecuen cia se llega a $\|v\|^2$, como se comprueba fácilmente haciendo a $\tau = 0$ en la ecuació (18*b*), es decir,

$$\int_{-\infty}^{\infty} G_v(f)\,df = R_v(0) = \|v\|^2 \qquad (20$$

Para interpretar esto y con ello justificar el nombre de $G_v(f)$ recordemos que $\|v\|^2$ e la energía o la potencia asociada con $v(t)$. Por lo tanto, se puede argüir que $G_v(f$ nos dice cómo la energía o la potencia se distribuyen en el dominio de la frecuencia merece ser designada como densidad espectral.

Para apoyar este punto de vista, sea $v(t)$ una señal de energía y $V(f)$ su es pectro. De la ecuación (16), $R_v(\tau) = [v(\tau)]*[v^*(-\tau)]$; así, recurriendo al teorem de la convolución,

$$G_v(f) = \mathscr{F}[R_v(\tau)] = \mathscr{F}[v(\tau)]\mathscr{F}[v^*(-\tau)]$$

De manera clara, $\mathscr{F}[v(\tau)] = \mathscr{F}[v(t)] = V(f)$, y es un ejercicio de rutina demostra que $\mathscr{F}[v^*(-\tau)] = \{\mathscr{F}[v(t)]\}^* = V^*(f)$. Por lo tanto,

$$G_v(f) = V(f)V^*(f) = |V(f)|^2 \qquad (21$$

a la cual se designa en la sección 2.3 como la *densidad espectral de potencia*. Invi tiendo la ecuación 21*a*), se tiene

$$R_v(\tau) = \mathscr{F}^{-1}[|V(f)|^2] = \int_{-\infty}^{\infty} |V(f)|^2 e^{j2\pi f\tau}\,df \qquad (21$$

que es una expresión alterna de la función de autocorrelación de una señal d energía. Más aún, con $\tau = 0$,

$$\|v\|^2 = \int_{-\infty}^{\infty} |V(f)|^2\,df \qquad (21$$

esto, en forma retrospectiva, es precisamente el *teorema de la energía de Rayleig* El lector puede comprobar la consistencia de la ecuación (21*b*), con la ecuación (17

Ahora considérese una señal periódica de potencia, expresada en forma de serie de Fourier

$$v(t) = \sum_{n=-\infty}^{\infty} c_v(nf_0)e^{j2\pi nf_0t} \qquad f_0 = \frac{1}{T_0}$$

Con ayuda de la ecuación (14) y del hecho de que todos los términos de la serie son mutuamente incoherentes,

$$R_v(\tau) = \sum_{n=-\infty}^{\infty} |c_v(nf_0)|^2 e^{j2\pi nf_0\tau} \qquad (22a)$$

y la transformación de Fourier da

$$G_v(f) = \sum_{n=-\infty}^{\infty} |c_v(nf_0)|^2\,\delta(f - nf_0) \qquad (22b)$$

Pero, ¿es esto la densidad espectral de potencia de una señal periódica? La respuesta es claramente afirmativa, puesto que una señal periódica se puede descomponer en términos de la forma $C_v(nf_0)e^{j2\Pi nf_0t}$, cada uno de los cuales representa una magnitud de potencia igual a $c_v(nf_0)^2$, localizada exactamente en la frecuencia $f = nf_0$. La densidad espectral es impulsiva en este caso nada más, pues se requiere de impulsos para indicar que hay unidades concentradas de potencia diferentes de cero, a frecuencias discretas. Por último, integrando la ecuación (22*b*) sobre todos los valores de f —o simplemente haciendo $\tau = 0$ en la ecuación (22*a*)— se llega al *teorema de la potencia de Parseval*.

Para resumir, mientras no hayamos comprobado de manera explícita que $G_v(f)$ representa la densidad espectral para otros tipos de señales, habremos presentado una fuerte evidencia en favor de esa conclusión basada en dos casos donde existe una firme noción intuitiva de cómo debe ser la densidad espectral. En conclusión, las funciones de densidad espectral se emplean con las funciones de transferencia para determinar las relaciones de entrada y salida. Pero antes de hacer esto, se debe señalar que $G_v(f)$, como $R_v(\tau)$, no representan en forma única a $v(t)$. En realidad, una señal dada tiene sólo una función de densidad espectral; no obstante, esa función de densidad espectral se puede aplicar a otras señales. Las sinusoides desplazadas en fase y los pulsos rectangulares retardados en el tiempo, son ejemplos sencillos. Y en el caso de señales aleatorias, dos o más formas de onda drásticamente diferentes pueden tener la misma densidad espectral —significando que sus promedios son los mismos, aunque las formas de onda sean diferentes.

Relaciones de entrada y salida

Respecto a la figura 2.38, sea $x(t)$ la entrada de un sistema lineal invariante en el tiempo, con una respuesta de impulso $h(t)$ y una función de transferencia $H(f)$. Si $x(t)$ es una señal de energía con espectro $X(f)$, la densidad espectral de salida de la energía es

$$|Y(f)|^2 = |H(f)|^2\,|X(f)|^2$$

o

$$G_y(f) = |H(f)|^2\,G_x(f) \qquad (23)$$

puesto que $G_x(f) = |X(f)|^2$ es la densidad espectral de energía de la entrada, etc.

Entrada — Sistema — Salida

$x(t)$ → $h(t)$ → $y(t)$

$G_x(f)$ → $H(f)$ → $G_y(f) = |H(f)|^2 G_x(f)$

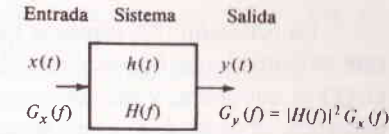



FIGURA 2.38

Aunque dedujimos la ecuación (23) para señales de energía, su forma sugiere que $H(f)^2$ siempre relaciona a las funciones de densidad espectral de entrada y salida, independientemente del tipo de la señal. Esto en realidad es correcto; la ecuación (23) se aplica para cualquier tipo de entrada cuya función de densidad espectral exista. La demostración general es relativamente directa, pero tediosa, y aquí sólo será bosquejada.

Primero, la autocorrelación $R_y(\tau)$ de la salida se encuentra insertando

$$y(t) = \int_{-\infty}^{\infty} h(\lambda)x(t-\lambda)\,d\lambda$$

en $R_y(\tau) = \langle y(t), y(t-\tau)\rangle$. Desarrollado esto, nos conduce a la pavorosa expresión

$$R_y(\tau) = \iint_{-\infty}^{\infty} h(\lambda)h^*(\mu)R_x(\tau + \mu - \lambda)\,d\mu\,d\lambda \qquad (24)$$

donde μ es otra variable fantasma. Entonces, tomando la transformada de Fourier, obtenemos finalmente

$$G_y(f) = \left[\int_{-\infty}^{\infty} h(\lambda)e^{-j2\pi f\lambda}\,d\lambda\right]\left[\int_{-\infty}^{\infty} h^*(\mu)e^{+j2\pi f\mu}\,d\mu\right]G_x(f)$$

en que $G_y(f) = H(f)H^*(f)G_x(f)$, como se aseguró.

Para reiterar lo significativo de este resultado, dada una señal de entrada de casi cualquier tipo, la función de densidad espectral correspondiente en la salida de un sistema lineal se encuentra multiplicando la densidad espectral de entrada por $H(f)^2$. Habiendo obtenido así $G_y(f)$, la aplicación de la ecuación (20) da la *energía* o la *potencia* de la señal de salida conforme

$$\|y\|^2 = \int_{-\infty}^{\infty} G_y(f)\,df = \int_{-\infty}^{\infty} |H(f)|^2 G_x(f)\,df \qquad (25)$$

A menudo, especialmente en aquellos problemas que incluyen señales aleatorias, $\|y\|^2$ es nada menos que la información buscada —y la ecuación (25) ofrece el itinerario más directo a esa información—. Se puede rebuscar alguna otra información relacionada con $y(t)$, por medio de

$$R_y(\tau) = \mathscr{F}^{-1}[G_y(f)] = \int_{-\infty}^{\infty} |H(f)|^2 G_x(f)e^{j2\pi f\tau}\,df \qquad (26)$$

Nótese, a propósito, que con esta transformada inversa es más fácil tratar que con la ecuación (24).

La ecuación (23) también hace expedito el cálculo de la densidad espectral, sea que se consideren filtros o no. Supongamos, por ejemplo, que $w(t) = dv(t)/dt$, $G_v(f)$ es conocida, y deseamos encontrar $G_w(f)$. De manera conceptual, $w(t)$ sería generada al pasar $v(t)$ a través de un diferenciador ideal, para el cual $H(f) = j2\pi f$ véase la ecuación (24), sección 2.3. Por lo tanto, empleando la ecuación (23), si

$$w(t) = \frac{dv(t)}{dt} \qquad (27a)$$

entonces

$$G_w(f) = (2\pi f)^2 G_v(f) \qquad (27b)$$

Asimismo, si

$$w(t) = \int_{-\infty}^{t} v(\lambda)\, d\lambda \qquad (28a)$$

entonces

$$G_w(f) = (2\pi f)^{-2} G_v(f) \qquad (28b)$$

Más adelante, en el siguiente capítulo, se presentan ilustraciones del empleo de las ecuaciones (23), (25) y (26), como se aplican a señales aleatorias.

2.7 PROBLEMAS

2.1 (Secc. 2.1) Emplee la ecuación (9) para encontrar $y(t)$ cuando $x(t) = 4 \cos 2\pi 10t$ y $H(j\omega) = 15 + j(\omega/\pi)$.
Respuesta: $100 \cos (2\pi 10t + 53°)$.

2.2 (Secc. 2.1) Encuentre y haga la gráfica de $|H(f)|$ y $[H(f)]$ para cada una de las siguientes funciones de transferencia:
(*a*) $(10 + jf)/(1 + jf)$
(*b*) $(1 + jf)/(10 + jf)$
(*c*) $(1 - jf)/(1 + jf)$

2.3 (Secc. 2.1) Si el capacitor en la figura 2.6*a* se reemplaza con un inductor, demuestre que $H(f) = j(f/B)/[1 + j(f/B)]$ donde $B = R/2\pi L$. Haga la gráfica de relación de amplitud y el corrimiento de fase.

2.4 (Secc. 2.1) En relación a la figura 2.6*a*, supóngase $x(t) = 10 \cos 2\pi f_0 t$ y el filtro tiene $B = 3$ kHz. Si $A_v = 2$, ¿cual es el valor de f_0?

2.5 (Secc. 2.2) Encuentre $\langle v(t)\rangle$ y P en $v(t) = Ae^{j(\omega_0 t + \theta)}$.
Respuesta: $0, A^2$.

2.6 (Secc. 2.2) Encuentre $\langle v(t)\rangle$ y P para la sinusoide totalmente rectificada $v(t) = A$ sen $2\pi t/T_0$.

2.7 (Secc. 2.2) Cuando $v(-t) = v(t)$, demuestre que

$$c(nf_0) = \frac{2}{T_0}\int_0^{T_0/2} v(t) \cos 2\pi n f_0 t \, dt$$

y emplee esto para demostrar la ecuación (11). Desarrolle un análisis similar para $v(-t) = -v(t)$.

2.8 (Secc. 2.2) Calcule $c(nf_0)$ para una onda triangular periódica con $v(t) = A(1 - 4|t|/T_0)$, $|t| \leq T_0/2$. (*Sugerencia*: véase el problema 2.7).

2.9 (Secc. 2.2) Calcule $c(nf_0)$ *para la onda totalmente rectificada* del problema 2.6.
Respuesta: $c(nf_0) = 2A/\pi(1 - n^2)n$ par; $c(nf_0) = 0$, n impar.

2.10 (Secc. 2.2) Emplee las ecuaciones (16) y (17) para comparar el espectro de una onda cuadrada con un tren de pulsos rectangulares en que $f_0\tau = ½$.

2.11 (Secc. 2.2) Supóngase que en $|f| > 1/\tau$ se remueven todos los componentes de frecuencia del espectro de un tren de pulsos rectangulares. Emplee el teorema de Parseval para calcular el porcentaje de potencia residual cuando $f_0\tau = ½$ y cuando $f_0\tau = 1/5$.

2.12 (Secc. 2.2) Con relación al ejemplo 2.3 calcule la amplitud de la componente sinusoide de salida en $2f_0$ cuando $N = 3$, $f_r = 3F_0$ y $Q = 12$. Compare con la amplitud de la componente en $3f_0$.

2.13 (Secc. 2.2) Una onda totalmente rectificada (problema 2.9) con $f_0 = 60$ Hz es la entrada de un filtro pasabajas RC. Encuentre el valor de B tal que la amplitud de la componente sinusoide mayor de salida es un 10% de la componente c.d.
Respuesta: 18.2 Hz.

2.14 (Secc. 2.3) Encuentre $V(f)$ para el *pulso gaussiano* $V(t) = Ae^{-\pi(t/\tau)^2}$, y verifique el efecto de extensión recíproca con gráficas de $v(t)$ y $V(f)$ para $\tau = 2$ y $\tau = 1$.
Respuesta: $A\tau e^{-\pi(f\tau)^2}$.

2.15 (Secc. 2.3) Encuentre $V(f)$ para el *pulso cosenoidal* $V(t) = A \cos \pi t/\tau$, $|t| \leq \tau/2$, dando su respuesta en términos de la función senc. Haga una gráfica de $V(f)$.

2.16 (Secc. 2.3) Demuestre la ecuación (13) sustituyendo a $v(t)$ con la expresión integral para $\mathcal{F}^{-1}[V(f)]$.

2.17 (Secc. 2.3) Emplee el teorema de Rayleigh para calcular el porcentaje de la energía total de un pulso exponencial (ejemplo 2.5) que está contenido en $|f| \leq 1/2T$ y $|f| \leq 1/T$.

2.18 (Secc. 2.3) Empleando los teoremas de la superposición y del retardo en el tiempo, encuentre las transformadas de las señales que se muestran en la figura P2.1. Construya gráficas de los espectro de amplitud eligiendo $\tau \ll T$.

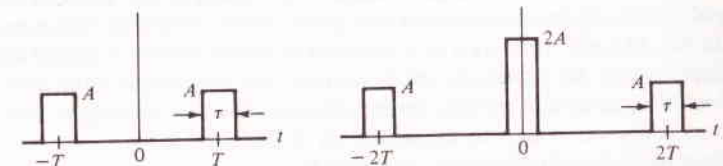



FIGURA P2.1

2.19 (Secc. 2.3) Encuentre $Z(f)$ en términos de $V(f)$ cuando $z(t) = v(at - t_d)$.
Respuesta: $|a|^{-1} V(f/a) \times e^{-j\omega t_d/a}$.

2.20 (Secc, 2.3) Emplee el teorema de la dualidad para encontrar la transformada de $A/[1 + (at)^2]$. (*Sugerencia*: véase la Tabla A.)

2.21 (Secc. 2.3) Genérese un nuevo par transformado aplicando el teorema de la dualidad al ejemplo 2.8.

2.22 (Secc. 2.3) Demuestre el teorema de la traslación de frecuencia, ecuación (21).

2.23 (Secc. 2.3) Emplee el teorema de la modulación para encontrar la transformada de $Ae^{-t/T}u(t)$ senc $\omega_c t$.

2.24 (Secc. 2.3) Emplee el teorema de la diferenciación y el problema 2.14 para encontrar la transformada de $Ate^{-\pi(t/\tau)^2}$.

2.25 (Secc. 2.4) Encontrar $v * w(t)$ cuando $v(t) = e^{-t}u(t)$ y $w(t) = \Pi(t/2T)$.
Respuesta: 0 para $t < -T$, $1 - e^{-(t+T)}$ para $-T < t < T$, $e^{-(t-T)} - e^{-(t-T)}$ para $t > T$.

2.26 (Secc. 2.4) Encuentre $v * v * v$ cuando $v(t) = \Pi(t/\tau)$.

2.27 (Secc. 2.4) Si $v_1(t) \longleftrightarrow V_1(f)$, etc., encuéntrese $V(f)$ para:
(*a*) $v(t) = [v_1(t) + v_2(t)] * v_3(t)$
(*b*) $v(t) = [v_1(t)v_2(t)] * v_3(t)$
(*c*) $v(t) = [v_1 * v_2(t)]v_3(t)$

2.28 (Secc. 2.4) Emplee el teorema de la convolución para encontrar la transformada del pulso trapezoidal de la figura 2.22*b*. Construya la gráfica del espectro de amplitud tomando $\tau_1 = 2\tau_2$.

2.29 (Secc. 2.4) Demuestre la ecuación (5).

2.30 (Secc. 2.4) Evalúe o simplifique cada una de las expresiones siguientes:
(*a*) $\int_{-\infty}^{\infty} t^2\delta(t-3)dt$
(*b*) $[t^2] * [\delta(t-3)]$
(*c*) $t^2\delta(t-3)$
(*d*) $[t^2] * [\delta(3t)]$

2.31 (Secc. 2.4) Emplee la ecuación (21) y el teorema del retardo en el tiempo para encontrar $c_w(nf_0)$ en términos de $c_v(nf_0$ cuando $w(t) = v(t - t_d)$.

2.32 (Secc. 2.4) La señal de la figura 2.16 puede escribirse como la convolución de $A\Pi(t/\tau)$ con $\delta(t + \tau/2)$. Emplee esta aproximación para obtener $Z(f)$.

2.33 (Secc. 2.4) Emplee la aproximación sugerida en el problema 2.32 para resolver el problema 2.18.

2.34 (Secc. 2.4) El teorema de la diferenciación se puede escribir en la forma

$$V(f) = (j2\pi f)^{-n}\mathscr{F}\left[\frac{d^n v(t)}{dt^n}\right]$$

la cual algunas veces simplifica el cálculo de transformadas. De manera específica, si la derivada n-ésima de $v(t)$ está formada en su totalidad de funciones simples (impulsos, pulsos rectangulares, etc.), entonces $[d^n v/dt^n]$ se encuentra fácilmente y $V(f)$ a partir de ella. El mejor valor de n es la derivada de menor orden, en la cual aparece primero el impulso, y es útil hacer gráficas de $v(t)$, dv/dt, etc., ya que así se detectan las discontinuidades, las cuales resultan ser impulsos cuando se diferencian. Emplee este método para encontrar la transformada de $v(t) = 2At\Pi(t/\tau)$.
Respuesta: $(jA\tau/\pi f)(\cos \pi f\tau - \text{senc } f\tau)$.

2.35 (Secc. 2.4) Emplee el método del problema 2.34 para encontrar la transformada de $A[1 - (2t/\tau)^2]\Pi(t/\tau)$.

2.36 (Secc. 2.5) Un sistema dado tiene una respuesta al impulso $h(t)$ y por función de transferencia $H(f)$. Obtenga expresiones para $y(t)$ y $Y(f)$ cuando:
(*a*) $x(t) = A\delta(t - t_d)$
(*b*) $x(t) = A[\delta(t + t_d) - \delta(t - t_d)]$

2.37 (Secc. 2.5) Resuelva el problema 2.36 para el caso de $x(t) = Ah(t - t_d)$

2.38 (Secc. 2.5) La entrada de un filtro pasabajas RC es $x(t) = \text{senc } 2Wt$. Construya una gráfica de la relación de energías E_y/E_x como una función de B/W.

2.39 (Secc. 2.5) Un pulso rectangular de duración τ es la entrada a una sujeción de orden cero, ejemplo 2.13. Emplee la aproximación del ejemplo 2.12 para encontrar $y(t)$ cuando $\tau \ll T$, $\tau = T$, y $\tau \gg T$.

2.40 (Secc. 2.5) La señal $x(t) = \Pi(t/\tau) + \Pi(t/\tau)\cos \omega_c t$ con $\tau = 1\,000/f_c$ se aplica a un filtro pasabajas RC con $B = 200$ Hz. Emplee la aproximación del ejemplo 2.12 para encontrar $y(t)$ cuando $f_c = 1$ kHz, $f_c = 10$ kHz y $f_c = 1$ MHz.

2.41 (Secc. 2.5) Encuentre $H(f)$ y $h(t)$ para el sistema cuyo diagrama se muestra en la figura P2.2.
Respuesta: $h(t) = T\Lambda[(t - T)/T]$.

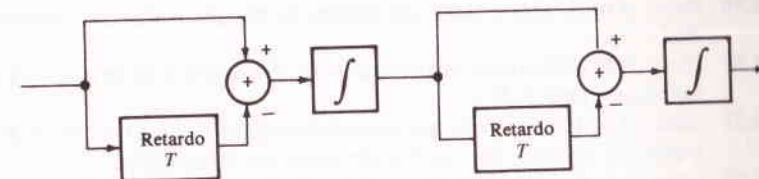



FIGURA **P2.2**

2.42 (Secc. 2.5) Cada uno de los cuadros marcado ZOH en la figura P2.3 son retenciones de orden cero. Encuentra $H(f)$ para el sistema completo.

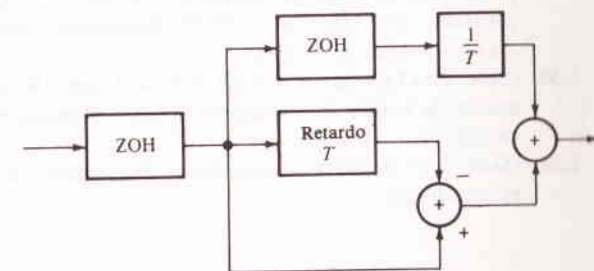



FIGURA **P2.3**

2.43 (Secc. 2.5) Defínase la *región de transición* de un filtro pasabajas como el intervalo de frecuencia positiva entre los anchos de banda a 3 y 10 dB. Encuentre esta región para un filtro pasabajas RC y exprésela como un porcentaje del ancho de banda a 3 dB.
Respuesta: 200%.

2.44 (Secc. 2.5) Repita el problema 2.43 para una filtro pasabajas *gaussiano* con $H(f) = K \exp[-(\ln 2/2)(f/B)^2]$, mostrando en primer lugar que B es el ancho de banda a 3 dB. ¿Es éste un filtro mejor que un filtro pasabajas RC?

2.45 (Secc. 2.5) La función de transferencia de un filtro pasaaltas *ideal* se puede escribir como $H(f) = Ke^{-j\omega t_d}[1 - \Pi(f/2f_{c0})]$ donde f_{c0} es la frecuencia de corte. Encuentre la respuesta al impulso e identifique la precursora.

2.46 (Secc. 2.5) ¿Cuántos elementos reactivos debe tener un filtro Butterworth (ejemplo 2.14), tales que el ancho de banda a 1 dB sea por lo menos de 0.9 B?

2.47 (Secc. 2.5) Construya la gráfica $H(f)$ para el filtro pasabajas de la figura P2.4. Haga la comparación con un Butterworth de segunda orden.

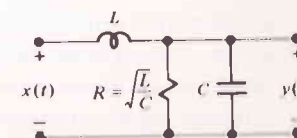



FIGURA **P2.4**

2.48 (Secc. 2.6) Dadas las señales de energía $v(t) = A\Pi(t/\tau)$ y $w(t) = Bt^2\Pi(t\tau)$, encuentre $\|v\|^2$, $\|w\|^2$, $\langle v(t),w(t)\rangle$, y compruebe la desigualdad de Schwartz.
Respuesta: $\|v\|^2 = A^2\tau$, $\|w\|^2 = B^2\tau^5/80$, $\langle v(t),w(t)\rangle = AB\tau^3/12$.

2.49 (Secc. 2.6) Por inspección, establezca las condiciones en α tales que cada uno de los siguientes pares de señales sean ortogonales:
(*a*) $\Pi(t/\tau)$, $\Pi[(t - \alpha)/\tau]$
(*b*) senc $2Wt$, senc $2Wt \cos 2\pi\alpha t$
(*c*) $e^{-|t|}$, t^α

2.50 (Secc. 2.6) Si $v(t)$ y $w(t)$ son señales de energía complejas, demostrar que $R_{wv}(\tau) = R^*_{vw}(-\tau)$.

2.51 (Secc. 2.6) Compruebe la afirmación en el ejemplo 2.15 de que $v(t)$ y $w(t - \tau)$ son ortogonales para toda τ.

2.52 (Secc. 2.6) Encuentre la autocorrelación de $A \cos(\omega_0 t + \theta) + B$. (*Sugerencia*: Demuéstrese primero que los dos términos son ortogonales.)

2.53 (Secc. 2.6) Emplee la ecuación (21) para encontrar $G_v(f)$ y $R_v(\tau)$ para las siguientes señales de energía:
(*a*) A senc $2Wt$
(*b*) $Ae^{-t/T}u(t)$

2.54 (Secc. 2.6) El pulso gaussiano $X(t) = 3e^{-\pi(10t)^2}$ (problema 2.14) se aplica a un filtro gaussiano con $H(f) = e^{-\pi(f/5)^2}$. Encuentre y haga la gráfica $R_x(\tau)$, $G_x(f)$, $G_y(f)$, y $R_y(\tau)$, y evalúe $\|y\|^2/\|x\|^2$.

2.55 (Secc. 2.6) La señal de energía $x(t) = A$ senc $2Wt$ se aplica a un filtro pasabajas RC con ancho de banda B. Encuentre $G_y(f)$ y obtenga expresiones aproximadas para $R_y(\tau)$ cuando $W \ll B$ y $W \gg B$.

2.56 (Secc. 2.6) Deduzca la ecuación (23) tomando la transformada de Fourier de $R_y(\tau)$, ecuación (24).

3

SEÑALES ALEATORIAS Y RUIDO

Las señales con las cuales se trató en el último capítulo fueron *determinísticas*, es decir, que escribiendo una función del tiempo explícita $v(t)$ se presume que el comportamiento de la señal es conocido y determinado durante todo el tiempo. Por otra parte, una característica de las *señales aleatorias* es que no se puede predecir su comportamiento de antemano. En este sentido y en forma particular, desde el punto de vista del extremo receptor, todas las señales de comunicación plenas de significado son aleatorias; porque si la señal fuera conocida de antemano no habría caso de transmitirla. Dicho de otra manera, una señal conocida no conduce información; ésta es una observación clave de la teoría de la información y será abordada nuevamente en el capítulo 9. Sin embargo, en éste se recalcará, en primer término, lo relativo a señales aleatorias *no deseadas o ruido*. Como el ruido siempre existe en todo sistema de comunicación, debe considerársele en toda descripción realista del funcionamiento del sistema.

Nuestro propósito es desarrollar los modelos o representaciones matemáticos de señales aleatorias para uso subsecuente. Y como tales señales no se pueden expresar como funciones explícitas del tiempo, se requiere de técnicas analíticas adicionales. En particular, cuando una señal aleatoria se examina sobre un largo periodo, puede tener ciertas regularidades que quedan descritas mejor en términos de probabilidades y promedios estadísticos. Así, a falta de una descripción exacta, hablare-

mos de valores promedio y de la probabilidad de que una señal aleatoria esté en un intervalo dado en un tiempo determinado.

Este capítulo empieza, por lo tanto, con los conceptos elementales de la teoría de las probabilidades. Se han introducido las variables aleatorias y su descripción estadística, recalcando la interpretación de señal aleatoria. Una vez obtenida una descripción estadística se relacionará al dominio de la frecuencia por medio de las funciones de autocorrelación y del teorema de Wiener-Kinchine. Por último, se examinan con cierto detalle las propiedades del ruido blanco y del ruido filtrado.

Conservando el alcance de las probabilidades y la estadística dentro de los objetivos del texto, se enfocarán de manera específica hacia aquellos aspectos que serán empleados en los capítulos subsecuentes. Además, hemos de confiar más en el razonamiento intuitivo que en la rigurosidad matemática. Quienes deseen profundizar sobre el tema, encontrarán bastante material en la literatura.†

3.1 INTRODUCCIÓN A LA PROBABILIDAD

La teoría de las probabilidades es una aproximación matemática a los fenómenos aleatorios. Como aleatorio significa impredecible, el conocimiento acerca de tal fenómeno requiere de una observación experimental de sus manifestaciones. En realidad, el lenguaje de la teoría de las probabilidades se expresa en base a experimentos y sus resultados; ésta es la forma aproximada que escogeremos para definir el concepto de probabilidades.

Considérese un experimento cuyo resultado varía de intento en intento y es de antemano desconocido; es decir, de un *experimento aleatorio*. Sea el evento A uno de los posibles resultados. Lanzar una moneda es un ejemplo, siendo los posibles resultados caras o cruces. El experimento se repite n veces y registra el número de veces que se presenta A, es decir, n_A. La relación n_A/n es la *frecuencia relativa de ocurrencia* del evento A para ese conjunto particular de intentos. Ahora, si el experimento tiene regularidad estadística, n_A/n se puede acercar a algún valor definido conforme n crece; es decir, incrementos mayores de n no ocasionan cambios significativos en n_A/n. Suponiendo que un valor límite existe, se define $P(A)$, la probabilidad del evento A, como la frecuencia relativa en el límite cuando $n \to \infty$,

$$P(A) \triangleq \lim_{n\to\infty} \frac{n_A}{n} \tag{1}$$

Así, la suposición de regularidad estadística nos lleva a nuestra definición de probabilidad. A partir de la definición, es claro que una probabilidad P es un número positivo limitado por

$$0 \leq P \leq 1 \tag{2}$$

† Por ejemplo, Papoulis (1965), Beckmann (1967), Drake (1967), Thomas (1969).

El hecho confirmado en forma experimental de que n_A/n se puede aproximar a una constante conforme aumenta el número de intentos, se conoce como *ley empírica de grandes números*. Aquí aparece la frase crucial "grandes números", pues sólo en el contexto de un gran número de repeticiones, la probabilidad resulta comprensible. Por ejemplo, si se lanza al aire una moneda 10 veces, puede salir cara 7 veces, $n_H/n = 0.7$. Pero experimentamos la certeza de que conforme n aumente, $n_H/n \to 0.5$, si la moneda no está "arreglada". Esta experiencia se expresa en la forma familiar de que "la posibilidad de que salga cara es 50%".

En muchos casos de lanzar al aire una moneda, es posible calcular las probabilidades sobre la base de condiciones de simetría, etc., sin ejecutar en realidad el experimento. O sea, que el experimento se hace en forma mental. Pero si la moneda es desviada por una acción irregular o intencional, las probabilidades requerirán, en efecto, una determinación empírica. Aparece un tercer caso cuando se desconoce la legalidad de la moneda y necesitamos estimar una probabilidad, digamos de caras, sin la ayuda de la experimentación. A falta de mayor información, la única suposición lógica sería $P(H) = ½$, puesto que no hay evidencia de que las caras sean más o menos probables que las cruces. (Esta aproximación mental algo simple es útil algunas veces, y ha recibido el nombre refinado de *el principio de la razón insuficiente*.)

EJERCICIO 3.1 Láncese al aire una moneda N veces (hasta quedar cansado) y hágase la gráfica n_H/n contra n

Eventos mutuamente exclusivos

Ahora considérese un experimento con m resultados *mutuamente exclusivos*, A_1, A_2, . . . , A_m, donde por mutuamente exclusivos se debe entender que la ocurrencia de cualquier resultado A_j no permite la ocurrencia simultánea de cualquier otro resultado. Surge entonces la pregunta: ¿Cuál es la probabilidad entre A_1 y A_2 en un intento dado? La probabilidad se escribe como $P(A_1 + A_2)$. Si $n_1 + n_2$ es el número de ocurrencias de A_1 o A_2 en n intentos, entonces

$$P(A_1 + A_2) = \lim_{n\to\infty} \frac{n_1 + n_2}{n} = P(A_1) + P(A_2)$$

donde $P(A_1) = \lim_{n\to\infty} (n_1/n)$, etc. Por extensión directa a dos o más resultados

$$P(A_i + A_j + A_k + \cdots) = P(A_i) + P(A_j) + P(A_k) + \cdots \tag{3}$$

y si se incluyen todos los resultados posibles

$$P(A_1 + A_2 + \cdots + A_m) = \sum_{j=1}^{m} P(A_j) = 1 \tag{4}$$

La ecuación (4) refleja simplemente que, en un intento dado, debe ocurrir uno de los resultados posibles. Si todos los resultados tienen las mismas probabilidades, se dice que son *igualmente probables*, entonces

$$P(A_j) = \frac{1}{m} \tag{5}$$

Se debe recalcar que las ecuaciones (3) a la (5) se aplican sólo a eventos mutuamente exclusivos.

En ocasiones interesará no la ocurrencia de un evento A, sino su *no ocurrencia*. Por conveniencia se designará A^c, léase "complemento de A", como el evento *no* A, similar a los términos correspondientes del álgebra booleana. Es obvio que A y A^c son mutuamente exclusivos, y en un intento dado, debe ocurrir uno u otro. Aplicando la ecuación (4), $P(A) + P(A^c) = 1$, o

$$P(A^c) = 1 - P(A) \tag{6}$$

lo cual es verdadero en general.

Ejemplo 3.1

Para ilustrar en algo estas ideas, supóngase que se lanzan dos monedas juntas. Hay cuatro resultados mutuamente exclusivos e igualmente probables, HH, HT, TH, TT (nótese que se considera que TH y HT son resultados diferentes). Puesto que el número de resultados igualmente probables es $m = 4$, la probabilidad de uno cualquiera es ¼. Así, $P(HH) = ¼$, y empleando complementos, $P(\text{no } HH) = 1 - ¼ = ¾$. Dos de los resultados dan una pareja repetida así $P(\text{pareja}) = P(HH) + P(TT) = ½$.

Probabilidades conjunta y condicional

En el ejemplo anterior, se habló de algunos eventos que no fueron idénticos a los resultados básicos del experimento, es decir, de "pareja" y "no HH". Esto constituye a menudo una práctica útil, pero entraña un peligro en el sentido de que los eventos que se derivan pueden no ser mutuamente exclusivos. Por ejemplo, los hechos derivados "no HH" y "pareja" pueden ocurrir de manera simultánea si el resultado es TT. Para tratar con eventos que no sean mutuamente exclusivos, se debe hablar en términos de *probabilidad conjunta*, es decir, la probabilidad de la ocurrencia conjunta de dos (o más) eventos.

Considérense los eventos A y B, que pueden o no ocurrir juntos, y sea su probabilidad conjunta $P(AB)$. Repítase el experimento n veces, donde n es muy grande, y regístrese n_{AB}, el número de veces que A y B ocurren juntos. La probabilidad conjunta es así

$$P(AB) = \frac{n_{AB}}{n}$$

En este proceso A ha ocurrido n_A veces, con o sin B, tal que n_A incluye n_{AB} y $n_A \geq n_{AB}$.

Es muy probable que la ocurrencia de B dependa en alguna forma de la ocurrencia de A. Para incluir tal información adicional, se define la *probabilidad condicional* $P(B/A)$, la probabilidad de B, en el supuesto que A haya ocurrido. De nuevo, aplicando la definición de frecuencia relativa a los n_A intentos en los que A ocurre,

$$P(B|A) = \frac{n_{AB}}{n_A} = \frac{n_{AB}/n}{n_A/n} = \frac{P(AB)}{P(A)} \tag{7a}$$

Puesto que n_{BA} es lo mismo que n_{AB}, invirtiendo el argumento se llega a

$$P(A|B) = \frac{P(AB)}{P(B)} \tag{7b}$$

Combinando las ecuaciones (7*a*) y (7*b*) se tienen dos relaciones para la probabilidad conjunta

$$P(AB) = P(B|A)P(A) = P(A|B)P(B) \tag{8}$$

Cuando se escribe la ecuación (8) en la forma

$$P(B|A) = \frac{P(B)P(A|B)}{P(A)}$$

que se conoce como el *teorema de Bayes*, el cual juega un papel importante en la teoría de la decisión estadística.

Eventos estadísticamente independientes

Supóngase que el evento B es independiente de A, tal que la ocurrencia de A no influye en la ocurrencia de B, y viceversa. En forma simbólica $P(B/A) = P(B)$ y $P(A/B) = P(A)$. Así, empleando la ecuación (8) se dice que dos eventos son *estadísticamente independientes* si

$$P(AB) = P(A)P(B) \tag{9}$$

Nótese en forma cuidadosa que la independencia estadística es muy diferente de la exclusividad mutua. De hecho, si A y B son mutuamente exclusivos, entonces, por definición, $P(AB) = 0$.

Es una cuestión simple entender o desarrollar la ecuación (9) a más de dos eventos. Así, si $A, B, C, \ldots$ son todos estadísticamente independientes, la probabilidad de su ocurrencia conjunta es

$$P(ABC\cdots) = P(A)P(B)P(C)\cdots \tag{10}$$

de la cual el lector debe encontrar la diferencia respecto de la ecuación (3).

Ejemplo 3.2

Continuando el ejemplo 3.1, desígnese el evento "no HH" con A y el "pareja" con B; $P(A) = ¾$ y $P(B) = ½$. El evento conjunto AB es "pareja y no HH", el cual debe ser TT. Por lo que $P(AB) = ¼$ y

$$P(B|A) = \frac{1/4}{3/4} = \frac{1}{3} \qquad P(A|B) = \frac{1/4}{1/2} = \frac{1}{2}$$

(¿Están estos resultados de acuerdo con los cálculos intuitivos?) El lector debe notar también que en este ejemplo $P(B/A) < P(B)$ mientras que $P(B/A^c) = 1 > P(B)$.

Es obvio que A y B no son estadísticamente independientes, puesto que

$$P(A)P(B) = 3/8 \neq P(AB)$$

Pero las monedas son en sí mismas estadísticamente independientes, si se lanzan en una forma propicia, y la ecuación (9) puede ser aplicada a los cálculos del ejemplo 3.1.

EJERCICIO 3.2 Una moneda es lanzada al aire tres veces. Calcule las siguientes probabilidades: $P(HTH)$, P(2 caras y una cruz), P(más caras que cruces | por lo menos una cruz). *Respuesta*: $(½)^3$, $3 \times (½)^3$, $4 \times (½)^3$, $3 \times (½)^3/7 \times (½)^3$.

3.2 VARIABLES ALEATORIAS Y FUNCIONES DE PROBABILIDAD

Los diversos juegos de probabilidad-suerte a cara o cruz, juego de naipes, etc., son temas naturales y fascinantes para el cálculo de probabilidades. Pero en el estudio de las comunicaciones, atañe lo referente a experimentos o procesos aleatorios en que se tienen resultados *numéricos*, por ejemplo, el valor de un voltaje de ruido en algún instante en el tiempo o el número de errores en un mensaje digital de datos. La descripción y análisis de tales *fenómenos aleatorios numéricamente valuados*, se facilitan en gran parte tratándolos como variables aleatorias y funciones de probabilidad.

Considérese un proceso o experimento aleatorio donde el resultado de cualquier observación se expresa como una cantidad numérica sencilla. La totalidad de *todos* los resultados posibles constituye un conjunto de números reales $x_1, x_2, \ldots, x_j, \ldots$, y los x_j pueden ser considerados como puntos sobre el eje real $-\infty < x < +\infty$. Los resultados posibles de este proceso son así limitados a un espacio unidimensional (es decir, una línea) conocido como el *espacio muestral*. Los diferentes puntos en el espacio muestral representan eventos mutuamente exclusivos, puesto que $x_i \neq x_j$.

Se asigna entonces al proceso una variable que representa el resultado de una observación dada y se le conoce como *variable aleatoria* o *variada*. La variable aleatoria X puede, por lo tanto, asumir cualquiera de los valores posibles del proceso con el cual está asociada. En otras palabras, X es el símbolo general para valores *observados*, mientras que x_j representa los valores *posibles*.

En este sentido, el concepto de variable aleatoria conduce a una útil notación taquigráfica, pues si el proceso tiene regularidad estadística, se le puede describir por medio de *funciones de probabilidad* definidas de manera suscinta en términos de X. Para ilustración, la probabilidad de que un valor observado sea mayor que x_k pero menor o igual que x_l se escribe $P(x_k < X \leq x_l)$, la cual es una función de x_k y de x_l —pero no X. Es curioso, pero las funciones de probabilidad que se definen empleando la aproximación de varible aleatoria *no son* funciones de la variable aleatoria.

Dependiendo de la naturaleza del proceso, una variable aleatoria puede asumir valores discretos *o* continuos. Si en cualquier intervalo finito del eje real, X puede asumir sólo un número finito de valores distintos, se conoce a la variable como aleatoria *discreta*. (Esta definición no excluye el caso de un número infinito de valores discretos, pero en un intervalo *finito* el número debe ser *finito*.) Por otra parte, si X puede tomar *cualquier* valor en un intervalo dado del eje real, se dice que es una variable aleatoria *continua*. Un proceso aleatorio a menudo se clasifica de acuerdo con el tipo de variada asociada a éste. Se estudiarán por separado ambos tipos, recalcando la continua, después de un breve análisis de funciones de distribución de probabilidad.

Funciones de distribución

La función de distribución acumulativa (CDF) de un proceso aleatorio se define como

$$F(x) \triangleq P(X \leq x) \tag{1}$$

$F(x)$ es simplemente la probabilidad de que un valor observado será menor o igual que la cantidad x y se aplica tanto a procesos continuos como discretos. Puesto que es una probabilidad, la función de distribución acumulativa está limitada por

$$0 \leq F(x) \leq 1 \tag{2}$$

Se obtienen otras propiedades de $F(x)$ de la siguiente consideración.

Para $x_1 < x_2$, los eventos $X \leq x_1$ y $x_1 < X \leq x_2$, son mutuamente exclusivos, y el evento $X \leq x_2$ es la combinación de los dos, es decir, $X \leq x_2$ implica que $X \leq x_1$ o $x_1 < X \leq x_2$. (Nótese la importancia de las desigualdades *abierta* y *cerrada* en los eventos que se especifican.) De la ecuación (3), sección 3.1, se tiene

$$P(X \leq x_2) = P(X \leq x_1) + P(x_1 < X \leq x_2)$$

o

$$P(x_1 < X \leq x_2) = F(x_2) - F(x_1) \tag{3}$$

Así, si se conoce la función de distribución para todos los valores de x, serán completamente conocidas las diferentes probabilidades asociadas con la variable aleatoria.

Puesto que las probabilidades son no negativas y la ecuación (3) es una probabilidad para cualquier $x_1 < x_2$, se concluye que la función de distribución acumulati-

va es siempre una *función no decreciente de* x. En forma analítica, $dF(x)/dx \geq 0$. Combinando esta observación con la ecuación (2) se tiene $F(-\infty) = 0$ y $F(+\infty) = 1$.

En lo que se refiere exclusivamente a probabilidades, la función de distribución acumulativa de un proceso aleatorio viene a ser una descripción completa y conveniente. Sin embargo, cuando se llega a los promedios estadísticos, $F(x)$ es menos útil que otras funciones de probabilidad que se consideren, ya sea que el proceso se presente del tipo discreto o continuo.

Variables aleatorias discretas. Funciones de la frecuencia

Una variable aleatoria discreta se puede describir por medio de la *función de la frecuencia*

$$P(x_j) \triangleq P(X = x_j)$$

donde x_j es uno de los valores discretos que puede asumir X y $P(X = x_j)$ es su probabilidad. La función de la frecuencia es, por lo tanto, una función de los valores posibles de x_j; el nombre procede de la interpretación de la frecuencia relativa de la probabilidad.

Al relacionar la función de la frecuencia con la función de distribución, se nota que las x_j son mutuamente exclusivas, así

$$F(x) = \sum_{x_j \leq x} P(x_j) \qquad (4)$$

Si hay m resultados posibles, enunciados todos, entonces

$$\sum_{j=1}^{m} P(x_j) = 1$$

es decir, que debe ocurrir uno de los resultados posibles en una observación dada.

Una inspección más a fondo de la ecuación (4) revela varios puntos interesantes con respecto a las funciones de distribución. Sean x_{k-1} y x_k dos resultados posibles *adyacentes*. Entonces

$$F(x_k) - F(x_{k-1}) = \sum_{j=1}^{k} P(x_j) - \sum_{j=1}^{k-1} P(x_j) = P(x_k)$$

lo cual es obvio a partir de la ecuación (3). De aquí se tiene entonces, que la función de distribución acumulativa de una variable aleatoria discreta tiene *discontinuidades en forma de escalón* en cada $x = x_j$, siendo la "altura" del escalón $P(x_j)$. Entre escalones, $F(x)$ es constante. En un punto de discontinuidad el valor de $F(x = x_j)$ se toma como el valor a la derecha de $x = x_j$, o sea, el valor mayor.

Ejemplo 3.3

Supóngase que se transmite un mensaje de datos formado de cuatro dígitos, sobre un sistema cuya probabilidad de error por dígito se conoce y es igual a 0.3. Considérese a X como el número *total* de errores observados en el mensaje recibido, de modo que los resultados posibles son $x_j = 0, 1, \ldots, 4$ errores. La función de la frecuencia $P(x_j)$ y los valores de $F(x)$ en $x = x_j$ se tabulan a continuación. En la siguiente sección se proporciona una técnica para obtener estas probabilidades; por

x_j	$P(x_j)$	$F(x_j)$
0	0.2401	0.2401
1	0.4116	0.6517
2	0.2646	0.9163
3	0.0756	0.9919
4	0.0081	1.0000

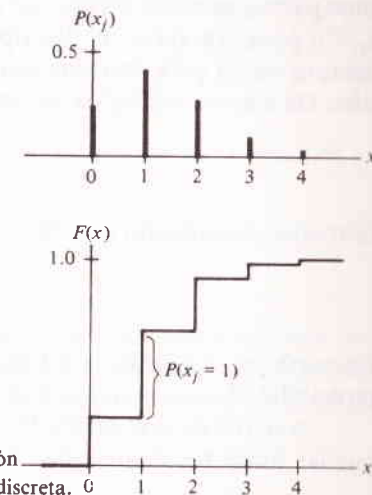



FIGURA 3.1
Función de la frecuencia y función de distribución acumulativa (CDF) para una variable aleatoria discreta.

el momento interesa $F(x)$ como se muestra en la figura 3.1 con $P(x_j)$. La figura demuestra con exactitud las propiedades de las funciones de distribución acumulativas; tanto las propiedades generales como las características de discontinuidad cuando la variada es discreta. Se recomienda examinar cuidadosamente la figura.

EJERCICIO 3.3 Supóngase que se lanzan al aire tres monedas y se designa con X el número total de caras que se obtienen. Calcúlese y constrúyase la gráfica $P(x_j)$ y $F(x)$.

Variables aleatorias continuas. Funciones de densidad

Para definir un poco el tema que nos ocupa, supóngase que se arrojan monedas en una línea dibujada en el piso, como en el popular pasatiempo escolar conocido como lanzamiento de centavos. La distancia entre la moneda y la línea es un número aleatorio que puede ser de cualquier valor, positivo o negativo, dentro de límites razonables, y se le puede asignar al proceso una variable aleatoria *continua* cuyo intervalo sea meramente esencial $[-\infty, +\infty]$.

En cualquier caso, el intervalo de la variada está limitado; en el continuo del intervalo existe un número infinito de valores posibles, y es clara la imposibilidad de

hacer distinción entre valores "adyacentes". (En contraste, los valores adyacentes de una variada *discreta* son identificables de manera unívoca; ésta es una definición asociada con la definición de la función de la frecuencia.) Por lo tanto, la descripción de variadas continuas en términos de función de la frecuencia, no es satisfactoria. En realidad, aun cuando X_1 sea un resultado posible en un proceso aleatorio continuo, $P(X = x_1)$ generalmente es cero, excepto ciertos casos muy especiales. El lector que cuestione esta afirmación debe tomar en cuenta la probabilidad de que la moneda lanzada caiga *exactamente* a 1 cm de la línea.

Aunque $P(X = x_1) = 0$ para cualquier valor de x_1, la moneda debe caer en alguna parte y se puede mencionar a tales eventos como $X \leq x_1$ o $x_1 < X \leq x_2$ si $x_1 < x_2$. En pocas palabras, la descripción de la función de distribución acumulativa es siempre válida para variadas continuas. Su derivada conduce a una mejor descripción. De manera específica, se define la *función de densidad de probabilidad* (PDF)

$$p(x) \triangleq \frac{dF(x)}{dx} \tag{5a}$$

Entonces, recordando que $F(-\infty) = 0$.

$$F(x) = \int_{-\infty}^{x} p(\lambda)\, d\lambda = P(X \leq x) \tag{5b}$$

En palabras: el *área* bajo la función de densidad de probabilidad de $-\infty$ a x es la probabilidad de que un valor observado será menor o igual que x.

A partir de la ecuación (5) y de las propiedades conocidas de $F(x)$, se aprecia que las funciones de densidad de probabilidad tienen las siguientes propiedades:†

$$p(x) = \frac{dF(x)}{dx} \geq 0$$

$$\int_{-\infty}^{\infty} p(x)\, dx = 1 \tag{6}$$

$$P(x_1 < X \leq x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} p(x)\, dx \tag{7}$$

Así, $p(x)$ es una función *no negativa* cuya área de x_1 a x_2 es simplemente la probabilidad de que x sea observada en este intervalo. Puesto que X debe encontrarse en algún lugar dentro de $[-\infty, \infty]$, el área total bajo $p(x)$ es la unidad.

Como un caso especial de la ecuación (7), sean $x_1 = x - dx$ y $x_2 = X$; la integral se reduce entonces a $p(x)dx$, así

$$p(x)\, dx = P(x - dx < X \leq x) \tag{8}$$

Esta ecuación proporciona otra interpretación muy útil de $p(x)$ y recalca su naturaleza como una *densidad* de probabilidad.

† Nótese que se debe emplear en la ecuación (5*b*) una variable fantasma λ —puesto que la integral es una función de x— pero no en las ecuaciones (6) y (7).

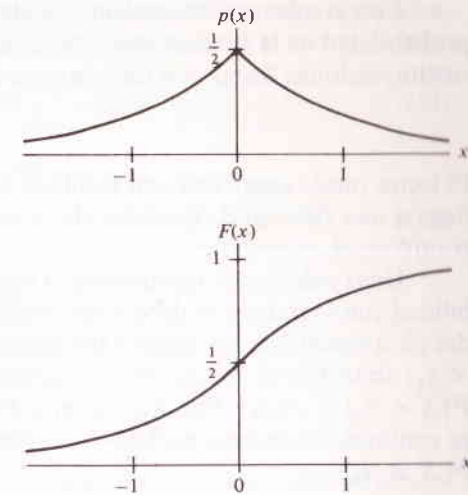



FIGURA 3.2
Función de densidad de probabilidad (PDF) y función de distribución acumulativa (CDF) para una variable aleatoria continua.

Ejemplo 3.4

Supóngase que la función de densidad de probabilidad de una variada continua es de la forma $p(x) = Ke^{-|x|}$, donde K es una constante no conocida. Para determinar K recurrimos a la propiedad del área unitaria de la ecuación (6); así, observando que $p(x)$ tiene simetría par,

$$\int_{-\infty}^{\infty} p(x)\, dx = 2\int_{0}^{\infty} Ke^{-x}\, dx = 2K = 1$$

Por lo tanto $K = \frac{1}{2}$ y

$$p(x) = \tfrac{1}{2}e^{-|x|} \tag{9a}$$

La función de densidad de probabilidad correspondiente es

$$F(x) = \begin{cases} \tfrac{1}{2}e^{x} & x \leq 0 \\ 1 - \tfrac{1}{2}e^{-x} & x > 0 \end{cases} \tag{9b}$$

que se obtiene empleando la ecuación (5*b*). La figura 3.2 muestra a $p(x)$ y $F(x)$. Ambas funciones son suaves y continuas, como deben ser cuando la variada es más bien continua que discreta.

Impulsos en funciones de densidad de probabilidad

Algunas veces una variada continua X tiene una componente discreta en el sentido de que un resultado valuado discreto particular, digamos x_0, para el cual $P(X = x_0) \neq 0$. En este caso $P(x)$ debe incluir un impulso, a saber, $P(x_0)\delta(x - x_0)$ y $F(x)$ presentará una discontinuidad de escalón de altura $P(x_0)$ en $x = x_0$.

Con el mismo razonamiento, podemos construir una función de densidad de probabilidad de la función de la frecuencia $P(x_j)$ de una variable aleatoria discreta sustituyendo las líneas con impulsos, es decir,

$$p(x) = \sum_j P(x_j)\delta(x - x_j) \tag{10}$$

El lector puede confirmar con facilidad que al aplicar las ecuaciones (5*b*) a la (10) se llega a una función de densidad de probabilidad en forma de escalón como se encontró en el caso discreto.

Unas palabras de advertencia al relacionar las funciones de densidad de probabilidad con impulsos: se debe tener cuidado especial cuando se emplean desigualdades para especificar los eventos (resultados). En la ecuación (10), por ejemplo, $P(X < x_1)$ no incluye el impulso en $x = x_1$ en tanto que $P(X \leq x_1)$ sí, pues $P(X \leq x_1) = P(X < x_1) + P(x_1)$. Por otra parte, si $P(x)$ no contiene impulsos —significa que X es continua de manera estricta— entonces no existe diferencia entre $P(X < x_1)$ y $P(X \leq x_1)$.

EJERCICIO 3.4 Sea la variable aleatoria Z, definida por

$$Z = \begin{cases} 0 & X \leq 0 \\ X & X > 0 \end{cases}$$

donde X tiene la función de densidad de probabilidad dada por la ecuación (9*a*) en el ejemplo 3.4. Demostrar que

$$p(z) = \tfrac{1}{2}\delta(z) + \tfrac{1}{2}e^{-z}u(z)$$

(*Sugerencia*: Primero bosquéjese $F(z)$ teniendo en cuenta que $P(Z < 0) = 0$.)

Transformación de variables aleatorias

El ejercicio anterior ha tocado ligeramente la transformación de variables aleatorias, donde una variable aleatoria se define como una función de otra y se desea encontrar la función de densidad de probabilidad resultante. Tales transformaciones pueden ser relativamente simples o muy difíciles, aunque aquí sólo se considerará una clase simple pero importante.†

Sea la variada continua Z una función singularmente valuada de X —cuya función de densidad de probabilidad $p_x(x)$ se conoce— y cuya función inversa existe, es decir,

$$Z = g(X) \qquad X = g^{-1}(Z)$$

Puesto que cada valor de Z corresponde a un valor *unívoco* de X, la probabilidad de encontrar a Z en algún intervalo diferencial dz es igual a la probabilidad de que X esté en el intervalo correspondiente dx. Así

$$p_z(z)\,|dz| = p_x(x)\,|dx| \tag{11}$$

† Para un tratamiento general, veáse Davenport y Root (1958, cap. 3).

donde se emplean los valores absolutos para asegurar que $p_z(z)$ será no negativa. Resolviendo para $p_z(z)$ e insertando $x = g^{-1}(z)$ se llega entonces a

$$p_z(z) = p_x(x)\left|\frac{dx}{dz}\right| = p_x[g^{-1}(z)]\left|\frac{dg^{-1}(z)}{dz}\right| \tag{12}$$

Para ilustración, dada la transformación lineal

$$Z = \alpha X + \beta \tag{13a}$$

se tiene $x = g^{-1}(z) = (z - \beta)/\alpha$ y

$$p_z(z) = \frac{1}{|\alpha|}\,p_x\!\left(\frac{z-\beta}{\alpha}\right) \tag{13b}$$

Funciones de densidad conjunta y condicional

Algunos procesos aleatorios se caracterizan por medio de dos o más variables aleatorias, las cuales pueden o no ser independientes, y un proceso con n variables aleatorias tiene un espacio muestral n-dimensional. La descripción de la función de densidad se extiende con rapidez en términos de funciones de densidad conjunta y condicional. Por su sencillez, se considera sólo el caso bidimensional.

Sean X y Y variadas continuas que se pueden observar en forma simultánea. Su función de densidad conjunta $p(x,y)$ se desarrolla conforme a la ecuación (8), a saber,

$$p(x,y)\,dx\,dy = P(x - dx < X \leq x, y - dy < Y \leq y) \tag{14}$$

(Nótese que una coma en el argumento se interpreta como conjunción "y".) Por lo tanto, la probabilidad de encontrar $-X$ en el intervalo $x_1 < X \leq x_2$ en tanto que al mismo tiempo $y_1 < Y \leq y_2$ es

$$P(x_1 < X \leq x_2, y_1 < Y \leq y_2) = \int_{y_1}^{y_2}\int_{x_1}^{x_2} p(x,y)\,dx\,dy$$

la cual es el equivalente bidimensional de la ecuación (7).

Si las dos variadas son *estadísticamente independientes*, entonces se puede demostrar que

$$p(x,y) = p_x(x)p_y(y) \tag{15}$$

Si no son independientes, necesitan expresarse las funciones de densidad condicional en la forma

$$p(x,y) = p_x(x)p(y\,|\,x) = p_y(y)p(x\,|\,y) \tag{16}$$

Por último, dada una función de densidad conjunta, la función de densidad de probabilidad para X sola se obtiene por el procedimiento siguiente. Si sólo interesa X, el valor de Y no importa, pues $P(X \leq x)$ es equivalente a $P(X \leq x, -\infty < Y \leq +\infty)$. Así

$$p_x(x) = \int_{-\infty}^{\infty} p(x,y)\, dy \tag{17}$$

EJERCICIO 3.5 Demuéstrese la ecuación (15) empezando con la ecuación (14) y aplicando la (8) de esta sección y la ecuación (9) de la sección 3.1.

3.3 PROMEDIOS ESTADÍSTICOS

Para muchos objetivos nuestros, la representación de la función de probabilidad de una variable aleatoria, proporciona mayor información de la que realmente es necesaria. De hecho, se puede probar que tal descripción completa es harto embarazosa, ocasionando más confusión que claridad. Así, a menudo se verá que es más sencillo y conveniente describir una variable aleatoria por medio de unos pocos *números* característicos que sintetizan, en forma breve, a la función de probabilidad. Estos números son los promedios estadísticos diferentes o valores medios.

Medias y momentos

El promedio estadístico o valor medio $\bar{x}$ de una variable aleatoria X es el promedio numérico de los valores que X puede adoptar ponderados por sus probabilidades. En forma esencial $\bar{x}$ es el promedio aritmético convencional de las observaciones, en el sentido de que para un número muy grande de intentos N, la *suma* de los valores observados se espera que sea aproximadamente $N_{\bar{x}}$. (De nuevo, la condición de grandes números resulta crucial.) Por esta razón, los valores medios se llaman también *valores esperados* o *expectativas* y se denotan como $\bar{x} = E[x]$.

Considérese un proceso aleatorio *discreto* en el cual se han hecho n observaciones. Si $X = x_1$ es observado n_1 veces, etc., la suma de los valores observados es precisamente

$$n_1 x_1 + n_2 x_2 + \cdots + n_m x_m = \sum_j n_j x_j$$

Dividiendo entre el número de intentos $n \to \infty$ se llega a

$$\bar{x} = \lim_{n\to\infty} \sum_j x_j \frac{n_j}{n} = \sum_j x_j P(x_j) \tag{1}$$

la cual se tomará como la definición de $\bar{x}$ para variadas discretas. Aplicando un razonamiento similar a variadas continuas, la suma se acerca a la integración de todos los valores de x; así

$$\bar{x} \triangleq \int_{-\infty}^{\infty} xp(x)\, dx \tag{2}$$

es la media de una variable aleatoria *continua*. En lo sucesivo todos los valores estadísticos se escribirán en términos de funciones de densidad, reconociendo que las variadas discretas quedan incluidas al escribir $p(x) = \Sigma P(x_j)\delta(x - x_j)$.

Si X es una variable aleatoria y $g(\ )$ es cualquier función analítica, entonces $g(X)$ es una variable aleatoria cuyo valor medio está dado en forma directa por

$$\mathbf{E}[g(x)] = \int_{-\infty}^{\infty} g(x)p(x)\, dx \tag{3}$$

Donde $g(X) = X^n$, su valor medio $\overline{x^n}$ se designa como el *momento n-ésimo* de X. El *primer momento* es, por supuesto, precisamente el valor medio $\bar{x}$. El *segundo momento* $\overline{x^2}$ es conocido comúnmente como el *valor cuadrático medio*. Obsérvese la diferencia entre el valor cuadrático medio y el valor medio cuadrado. El orden de las operaciones —promedio y cuadrado— no es intercambiable, por lo que $\overline{x^2} \neq \bar{x}^2$.

Los promedios en tiempo y estadísticos tienen ciertas similitudes; en particular, ambos constituyen operaciones *lineales*. Así, si α y β son constantes, entonces

$$\mathbf{E}[\alpha x + \beta] = \alpha\bar{x} + \beta \tag{4}$$

Como se demuestra con facilidad empleando la ecuación (3) con $g(x) = \alpha x + \beta$. Aunque este resultado parece más trivial, conduce a la relación no tan obvia

$$\mathbf{E}[\bar{x}x] = \bar{x}^2$$

donde $\bar{x}$ es una constante.

Variancia, desviación estándar y desigualdad de Chebyshev

Una forma ligeramente modificada del segundo momento recibe el nombre de *variancia*, definida como

$$\sigma_x^2 \triangleq \mathbf{E}[(x - \bar{x})^2] \tag{5a}$$

donde se debe advertir con cuidado el orden de las operaciones. La raíz cuadrada de la variancia, designada como *desviación estándar*, es una medida de la *anchura* de los valores observados. Antes de demostrar esta afirmación, obténgase una expresión más conveniente para σ_x^2 recurriendo a la linealidad del promedio estadístico:

$$\sigma_x^2 = \mathbf{E}[x^2 - 2x\bar{x} + \bar{x}^2] = \overline{x^2} - 2\bar{x}^2 + \bar{x}^2$$

o

$$\sigma_x^2 = \overline{x^2} - \bar{x}^2 \tag{5b}$$

Por lo que la variancia es simplemente el valor cuadrado medio menos el valor medio al cuadrado.

Para la interpretación de la variancia y de la desviación estándar, se empleará la *desigualdad de Chebyshev*†

$$P(|X - \bar{x}| > \kappa\sigma_x) < \frac{1}{\kappa^2} \qquad \kappa > 0 \tag{6a}$$

† Se pronuncia también Tchebycheff.

Así, por ejemplo,

$$\mathbf{E}[x + y] = \iint_{-\infty}^{\infty} (x + y)p(x,y)\, dx\, dy$$

$$= \int_{-\infty}^{\infty} x \underbrace{\left[\int_{-\infty}^{\infty} p(x,y)\, dy\right]}_{p_x(x)} dx + \int_{-\infty}^{\infty} y \underbrace{\left[\int_{-\infty}^{\infty} p(x,y)\, dx\right]}_{p_y(y)} dy$$

Por lo que,

$$\overline{x + y} = \int_{-\infty}^{\infty} x p_x(x)\, dx + \int_{-\infty}^{\infty} y p_y(y)\, dy = \bar{x} + \bar{y} \tag{8}$$

un resultado más simple y satisfactorio. Generalizando, se puede decir que la *media de una suma es la suma de los valores medios*, sin importar la dependencia estadística.

Se aplica una conclusión paralela al *producto* de dos o más variadas, advertidos de que son *estadísticamente independientes*. Así, si X y Y son variables aleatorias independientes, entonces

$$\overline{xy} = \bar{x}\bar{y} \tag{9}$$

Generalizando esta regla, si X y Y son independientes, entonces

$$\mathbf{E}[g_1(x)g_2(y)] = \mathbf{E}[g_1(x)]\mathbf{E}[g_2(y)] \tag{10}$$

donde g_1 y g_2 son funciones de x y y, respectivamente.

Ahora considérese $Z = X + Y$, donde X y Y son independientes. Claro que $\bar{z} = \bar{x} + \bar{y}$, pero ¿cuál es la variancia $\sigma_z{}^2$? Para contestar esto, primero calcúlese $\overline{z^2}$ por medio de

$$\overline{z^2} = \mathbf{E}[x^2 + 2xy + y^2] = \overline{x^2} + 2\bar{x}\bar{y} + \overline{y^2}$$

donde con el empleo de las ecuaciones (4), (8) y (9) se ha simplificado en mucho este paso. Por último, puesto que $\sigma_z{}^2 = \overline{z^2} - \bar{z}^2$ y $\bar{z}^2 = (\bar{x} + \bar{y})^2$, se tiene

$$\sigma_z{}^2 = (\overline{x^2} - \bar{x}^2) + (\overline{y^2} - \bar{y}^2) = \sigma_x{}^2 + \sigma_y{}^2 \tag{11}$$

En general, la variancia de una suma es la suma de las variancias, sin olvidar que los componentes son independientes.

EJERCICIO 3.7 Demuéstrese la ecuación (10) por medio de la inserción de la ecuación (15), sección 3.2, en la ecuación (7).

3.4 MODELOS ÚTILES DE PROBABILIDAD

Hay bastantes modelos de probabilidad, tanto discretos como continuos, algunos de los cuales se dan a continuación. Por fortuna, los pocos que se analizarán cubren muchos de los casos que se verán más adelante.

Distribución binomial

Considérese un experimento con dos resultados posibles A y $B = A^c$, el lanzar al aire una moneda sencilla, por ejemplo. Sean las probabilidades $P(A) = \epsilon$ y $P(B) = 1 - \epsilon$. Si el experimento se repite m veces y se totaliza el número de veces que A se repite, se habrá generado un proceso aleatorio discreto numéricamente valuado. Asignando la variable aleatoria N como el número de ocurrencias de A, entonces N puede tomar los $m + 1$ valores discretos $0, 1, \ldots, m$.

La función de la frecuencia correspondiente $P(N = n)$ se conoce como la *distribución binomial* que se escribe

$$P_m(n) = \binom{m}{n} \epsilon^n (1 - \epsilon)^{m-n} \tag{1}$$

donde

$$\binom{m}{n} \triangleq \frac{m!}{n!\,(m - n)!} \tag{2}$$

expresión que se conoce como *coeficiente binomial*, puesto que es igual al $(n + 1)$ésimo coeficiente en la expansión de $(x + y)^m$. Se ve fácilmente que

$$\binom{m}{m-n} = \binom{m}{n} \qquad \text{y} \qquad \binom{m}{1} = m$$

Aparecen algunos problemas cuando $n = 0$, pero esto se resuelve por medio de la definición $0! \triangleq 1$ de tal suerte que

$$\binom{m}{0} = \binom{m}{m} = 1$$

Otros valores se dan en tablas o se pueden encontrar por medio del triángulo de Pascal. Se omite aquí la deducción de la ecuación (1) en aras de brevedad.

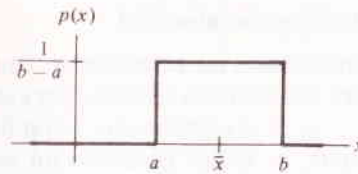
Mientras que el lanzamiento de monedas no es particularmente interesante para nosotros, la ocurrencia de errores en la transmisión digital de datos sí lo es en forma definitiva. Y si la probabilidad de error por dígito es la misma que para todos los dígitos, digamos ϵ, entonces $P_m(n)$ da la probabilidad de n errores en un mensaje de m dígitos, etc. Asimismo

$$F_m(k) = \sum_{n=0}^{k} P_m(n)$$

es la probabilidad de k o de menos errores en m dígitos. Como ejercicio, el lector quizá desee comprobar los valores de la función de la frecuencia y la función de distribución dados en el ejemplo 3.3, sección 3.2, donde $m = 4$ t y $\epsilon = 0.3$.

Volviendo a los promedios estadísticos de N, el valor medio y la variancia de la distribución binomial se encuentran insertando $P_m(n)$ en la ecuación (1), sección 3.3, etc. Se obtiene, después de algunos desarrollos

$$\bar{n} = m\epsilon \qquad \sigma_n{}^2 = m\epsilon(1 - \epsilon) = \bar{n}(1 - \epsilon) \tag{3}$$

FIGURA 3.3
Función de densidad de probabilidad uniforme.

y puesto que σ_n/n es proporcional a $1/\sqrt{m}$, la *anchura relativa* de los valores probables disminuye conforme aumenta m.

La ecuación (3) es de particular importancia para el cálculo rápido del número esperado de errores en un mensaje digital. Por ejemplo, si se envían 10 000 dígitos con $\epsilon = 0.01$, entonces $\bar{n} = 100$ y $\sigma_n^2 = 99$; tomando $\bar{n} \pm 2\sigma_n$ como el intervalo esperado, habrían unos 80 a 100 errores en un mensaje típico. Incidentalmente, esto no constituye una muy buena proporción de errores.

Distribución uniforme

Cuando es igualmente probable que una variada continua tome cualquier valor en el intervalo $[a, b]$, se dice que es *uniformemente distribuida*. La función de densidad de probabilidad es

$$p(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{de otra manera} \end{cases} \tag{4}$$

como se muestra en la figura 3.3. El valor medio, el valor cuadrado medio y la variancia son

$$\begin{aligned} \bar{x} &= \int_{-\infty}^{\infty} \frac{x}{b-a}\,dx = \frac{b+a}{2} \\ \overline{x^2} &= \int_{-\infty}^{\infty} \frac{x^2}{b-a}\,dx = \frac{b^2+ab+a^2}{3} \\ \sigma_x^2 &= \frac{(b-a)^2}{12} \end{aligned} \tag{5}$$

El valor medio cae justo donde se le espera, precisamente a la mitad del intervalo. En otras palabras, la variancia depende sólo del *tamaño* del intervalo, $b - a$, y no de su posición absoluta.

Distribución gaussiana

El modelo de probabilidad más ampliamente conocido es el de *distribución gaussiana* o *normal*. Esta función se presenta en muchas aplicaciones de análisis estadístico en razón de un interesante fenómeno conocido como el *teorema del límite central*, que para nuestro objetivo, se puede establecer de la siguiente manera:

FIGURA 3.4
Función de densidad de probabilidad gaussiana.

> Si $X_1, X_2, \ldots, X_M$ son variables aleatorias independientes y si se considera a la variable aleatoria Y como su suma, entonces conforme M crece adquiriendo valores muy grandes, la distribución de Y se aproxima a una distribución gaussiana. El resultado es independiente de las distribuciones de las componentes individuales, mientras que la contribución de cada una es pequeña comparada con su suma.

No se intentará la demostración del teorema, ni será usado en forma explícita en lo sucesivo, ya que se debe tener mucho cuidado en su aplicación. (Se ha dicho que el teorema del límite central es una herramienta peligrosa en manos de aficionados.) No obstante, se puede observar que el ruido eléctrico es a menudo la suma de los efectos de un gran número de electrones moviéndose en forma aleatoria y, por lo tanto, se tendrá una distribución gaussiana. De manera similar, los errores aleatorios en mediciones experimentales se deben a muchas causas irregulares y fluctuantes; tales errores harán que los valores medidos sean aleatorios con una distribución gaussiana respecto del valor verdadero.

La función de densidad de probabilidad de una variada de distribución gaussiana es la conocida curva en forma de campana

$$p(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-m)^2/2\sigma^2} \tag{6}$$

Como se vio en la figura 3.4, $p(x)$ describe a una variable aleatoria continua que puede adquirir cualquier valor en $[-\infty, +\infty]$ pero es más probable encontrarla cerca de $x = m$. En efecto, como el lector puede verificarlo, m es el valor medio de X. Y, como se desprende la notación, la desviación estándar es σ. También debe notarse la simetría par de $p(x)$ respecto de $x = m$. Una consecuencia inmediata es

$$P(X \le m) = P(X > m) = \tfrac{1}{2} \tag{7}$$

Supóngase entonces que un voltaje de ruido aleatorio tiene una distribución gaussiana y que se desea encontrar la probabilidad de que un valor observado caerá fuera del intervalo $m \pm \kappa\sigma$. Para simplificar un poco, la simetría de $p(x)$ implica $P(|X - m| > \kappa\sigma) = 2P(X > m + \kappa\sigma)$. Estableciendo esto como una integral,

$$P(X > m + \kappa\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{m+\kappa\sigma}^{\infty} e^{-(x-m)^2/2\sigma^2}\,dx$$

la cual establece que la probabilidad de encontrar a X fuera de las desviaciones estándar $\pm \kappa$ del valor medio es menor que $1/\kappa^2$, sin importar $p(x)$. Por ejemplo, con $\kappa = 2$, $P < 0.25$; con $\kappa = 10$, $P < 0.01$. La desigualdad de Chebyshev también se escribe en términos del evento $|X - \bar{x}| \leq \kappa\sigma_x$, el cual es el complemento del caso anterior, así

$$P(|X - \bar{x}| \leq \kappa\sigma_x) \geq 1 - \frac{1}{\kappa^2} \tag{6b}$$

Así, se puede especificar un intervalo, digamos $\bar{x} \pm 2\sigma_x$ dentro del cual se esperan valores observados con gran probabilidad, $P \leq 0.75$. Más aún, una *desviación estándar grande* implica una *anchura grande* de valores probables para alguna observación dada y viceversa. Cuando aplicamos esto a señales eléctricas aleatorias, se logra una mejor interpretación física de σ. La demostración de la ecuación (6) se puede encontrar en cualquier texto sobre probabilidad y estadística.

Ejemplo 3.5

A modo de ilustración, se aplican las ecuaciones anteriores al ejemplo 3.4, sección 3.2, donde $p(x) = \frac{1}{2}e^{-x}$. Puesto que $p(x)$ tiene simetría par

$$\bar{x} = \int_{-\infty}^{\infty} xp(x)\,dx = 0$$

$$\overline{x^2} = \int_{-\infty}^{\infty} x^2 p(x)\,dx = 2\int_0^{\infty} x^2 \tfrac{1}{2}e^{-x}\,dx = 2$$

y por lo tanto

$$\sigma_x^2 = 2 - 0 = 2 \qquad \sigma_x = \sqrt{2}$$

La probabilidad de que un valor observado caerá dentro de $\pm 2\sigma_x$ del valor medio $\bar{x} = 0$ es

$$P(|X| \leq 2\sqrt{2}) = \int_{-2\sqrt{2}}^{2\sqrt{2}} p(x)\,dx = 2\int_0^{2\sqrt{2}} \tfrac{1}{2}e^{-x}\,dx = 0.94$$

comparado con $P \geq 0.75$ de la desigualdad de Chebyshev. ////

EJERCICIO 3.6 Sea $Z = 3X + 5$ donde X es como en el ejemplo 3.5. Sin integrar, demuéstrese que $\bar{z} = 5$ y $\overline{z^2} = 43$. Después encuéntrese $p_z(z)$ empleando la ecuación (13*b*), sección 3.2, y compruébense estos valores por integración.

Variables aleatorias múltiples

En el caso de dos o más variables aleatorias, los valores medios se encuentran empleando la integración múltiple y las funciones de densidad conjunta en la forma general

$$\mathbf{E}[g(x,y)] = \iint_{-\infty}^{\infty} g(x,y)p(x,y)\,dx\,dy \tag{7}$$

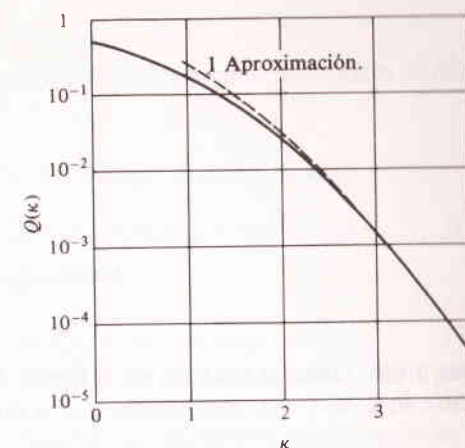



FIGURA 3.5

pero, por desgracia, no se puede resolver en la forma cerrada y requiere evaluación numérica. Por consiguiente, se introduce la función

$$Q(\kappa) \triangleq \frac{1}{\sqrt{2\pi}}\int_{\kappa}^{\infty} e^{-\lambda^2/2}\,d\lambda \tag{8}$$

la cual se ha tabulado en forma amplia.† Haciendo $\lambda = (x - m)/\sigma$, se ve que

$$P(X > m + \kappa\sigma) = P(X \leq m - \kappa\sigma) = Q(\kappa) \tag{9a}$$

y

$$P(|X - m| > \kappa\sigma) = 2Q(\kappa) \tag{9b}$$

A partir de la ecuación (9*a*), $Q(\kappa)$ se puede enunciar como el *área bajo la cola gaussiana*.

Cuando $\kappa \gg 1$, la probabilidad anterior es muy pequeña, tan pequeña como para hacer que muchas tablas de $Q(\kappa)$ sean inútiles para el cálculo. En este caso, hay una aproximación analítica que se obtiene de la ecuación (8) integrando por partes, a saber,

$$Q(\kappa) \approx \frac{1}{\sqrt{2\pi}\,\kappa} e^{-\kappa^2/2} \qquad \kappa \gg 1 \tag{10}$$

La figura 3.5 es una gráfica de $Q(\kappa)$ donde se muestra el valor exacto y la aproximación. Se ve que lo último es bastante exacto para $\kappa \geq 3$. Ahora considérese la probabilidad $P(m < X \leq m + \kappa\sigma)$. De nuevo resultan útiles la simetría y la interpretación de área de la función de densidad, pues se puede dividir $p(x)$ en

† La tabla D muestra una gráfica detallada de $Q(k)$ así como su relación con otras funciones definidas por varios autores.

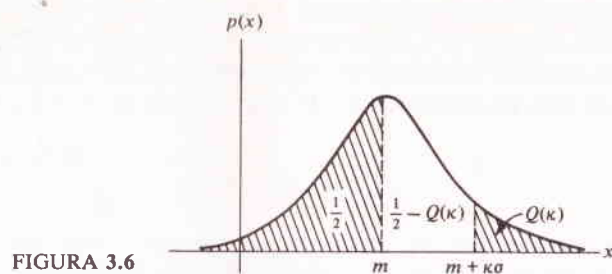



FIGURA 3.6

tres áreas, como se muestra en la figura 3.6. La probabilidad en cuestión es el área entre m y $m + \kappa\sigma$, recordando que el área total es la unidad,

$$P(m < X \leq m + \kappa\sigma) = 1 - \tfrac{1}{2} - Q(\kappa)$$

así,

$$P(m < X \leq m + \kappa\sigma) = P(m - \kappa\sigma < X \leq m) = \tfrac{1}{2} - Q(\kappa) \tag{11a}$$

y

$$P(|X - m| \leq \kappa\sigma) = 1 - 2Q(\kappa) \tag{11b}$$

Lo último también proviene en forma directa de la ecuación (9b).

Por último, empleando la tabla 3.1, se comparan algunos valores exactos de $P(X - m \leq \kappa\sigma)$ con la desigualdad de Chebyshev, $P(X - \bar{x} \leq \kappa\sigma) \geq 1 - 1/\kappa^2$. De manera obvia, el límite inferior $1 - 1/\kappa^2$ es conservativo. Más aún, el intervalo esperado de los valores observados es algo menor que $m \pm 2\sigma$; un intervalo más realista y empleado con frecuencia es $m \pm \sigma$, para el cual la probabilidad es 0.68.

EJERCICIO 3.8 Si X es gaussiana con $\bar{x} = 5$ y $\overline{x^2} = 89$, demuéstrese que $P(9 < X \leq 25) = Q(0.5) - Q(2.5) \approx 0.30$. (*Sugerencia*: Constrúyase la gráfica de función de densidad de probabilidad y señálense los límites del área en cuestión en términos de $m + \kappa\sigma$.)

Tabla 3.1

κ	$1 - 2Q(\kappa)$	$1 - \dfrac{1}{\kappa^2}$
0.5	0.38	
1.0	0.68	0.00
1.5	0.87	0.56
2.0	0.95	0.75
2.5	0.99	0.84

Distribución gaussiana bivariada

Dos variables aleatorias X_1 y X_2 son conjuntamente gaussianas si su función de densidad conjunta es

$$p(x_1,x_2) = \frac{1}{2\pi\sqrt{C_{12}}} \exp\{-[\sigma_2{}^2(x_1 - m_1)^2 + \sigma_1{}^2(x_2 - m_2)^2 - 2\sigma_{12}(x_1 - m_1)(x_2 - m_2)]/2C_{12}\} \tag{12}$$

donde $m_1 = \overline{x_1}$, $\sigma_1{}^2 = \overline{x_1{}^2} - m_1{}^2$, etc., y

$$\sigma_{12} \triangleq E[(x_1 - m_1)(x_2 - m_2)] = \overline{x_1 x_2} - m_1 m_2 \tag{13a}$$

$$C_{12} \triangleq \sigma_1{}^2\sigma_2{}^2 - \sigma_{12}{}^2 \tag{13b}$$

El parámetro σ_{12} se conoce como la *covariancia*; un parámetro relacionado es el *coeficiente de correlación*

$$\rho_{12} \triangleq \frac{\sigma_{12}}{\sigma_1\sigma_2} \tag{13c}$$

Nótese que $C_{12} = \sigma_1{}^2\sigma_2{}^2(1 - \rho_{12}{}^2)$.

Aunque la ecuación (12) parece ser una expresión más que formidable, simplemente representa una superficie en forma de campana sobre el plano x_1x_2 cuyo máximo se encuentra en $x_1 = m_1$, $x_2 = m_2$. Esta observación, con algunos cálculos de rutina, revela que todas las funciones de densidad individuales y condicionales son también gaussianas. Así

$$p_{x_1}(x_1) = \int_{-\infty}^{\infty} p(x_1,x_2)\,dx_2 = \frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-(x_1 - m_1)^2/2\sigma_1{}^2} \tag{14a}$$

$$p(x_1|x_2) = \frac{p(x_1,x_2)}{p_{x_2}(x_2)} = \frac{1}{\sqrt{2\pi(1 - \rho_{12}{}^2)}\sigma_1} e^{-[(x_1 - m_1) - \rho_{12}(\sigma_1/\sigma_2)(x_2 - m_2)]^2/2(1 - \rho_{12}{}^2)\sigma_1{}^2} \tag{14b}$$

e intercambiando los índices se tiene $p_{x_2}(x_2)$, etc. Si las variables son *estadísticamente independientes*, entonces $\sigma_{12} = 0$ y la ecuación (12) se reduce a

$$p(x_1,x_2) = p_{x_1}(x_1)p_{x_2}(x_2)$$

como se esperaba. Esto procede también de que $\rho_{12} = 0$ y $p(x_1|x_2) = p_{x_1}(x_1)$. De manera incidental, las variadas gaussianas *no correlacionadas* son también *independientes*, pero esto no es necesariamente válido para otros modelos de probabilidad; en el problema 3.28 se ve un ejemplo.

Ahora considérese una nueva variable aleatoria Z constituida por combinaciones lineales de X_1 y X_2, es decir,

$$Z = \alpha_1 X_1 + \alpha_2 X_2$$

donde α_1 y α_2 son constantes. Se puede demostrar que

$$p_z(z) = \frac{1}{\sqrt{2\pi}\,\sigma_z} e^{-(z-m_z)^2/2\sigma_z^2} \tag{15}$$

con

$$m_z = \alpha_1 m_1 + \alpha_2 m_2$$
$$\sigma_z^2 = \alpha_1^2\sigma_1^2 + \alpha_2^2\sigma_2^2 + 2\alpha_1\alpha_2\rho_{12}\sigma_1\sigma_2$$

Por lo tanto, *una combinación lineal de variadas gaussianas conjuntas es también una variada gaussiana*. De nuevo ésta es una característica unívoca de las variables aleatorias gaussianas.

El análisis anterior se puede extender a más de dos variadas gaussianas. No obstante, la notación se vuelve intratable, a menos que se empleen símbolos matriciales.

3.5 SEÑALES ALEATORIAS

Bajo este punto de vista, se ha hablado acerca de las variables aleatorias y sus promedios estadísticos, en términos muy generales. Ahora se concentrará la atención en las *señales aleatorias*, es decir, las manifestaciones de *procesos eléctricos aleatorios* que ocurren en el *tiempo*, designados también como *procesos estocásticos*. Así, si $v(t)$ es una señal aleatoria, la variable aleatoria V puede representar los valores de $v(t)$ en los tiempos de observación.

Bajo ciertas condiciones, se denomina *ergodicidad*, a una relación comprensible de manera intuitiva, entre los promedios estadísticos y los promedios de tiempo que serán obtenidos. Esto a su vez conduce a la descripción espectral de las señales aleatorias por medio de funciones de autocorrelación y del teorema de Wiener-Kinchine.

Se da principio considerando los procesos ergódicos y las implicaciones de la ergodicidad.

Procesos ergódicos. Promedios de ensamble

Sea $v(t)$ la forma de onda de voltaje producida por algún generador de ruido, de tal manera que $v(t)$ es una señal aleatoria de potencia promedio finita. Si se examina la forma de onda durante un periodo largo, se pueden medir o calcular sus diferentes promedios *de tiempo*, a partir de las definiciones del capítulo 2. Además, el significado de estos promedios es aparente: $\langle v(t)\rangle$ es la componente de CD del ruido $\langle v^2(t)\rangle$ es su potencia promedio, y así sucesivamente.

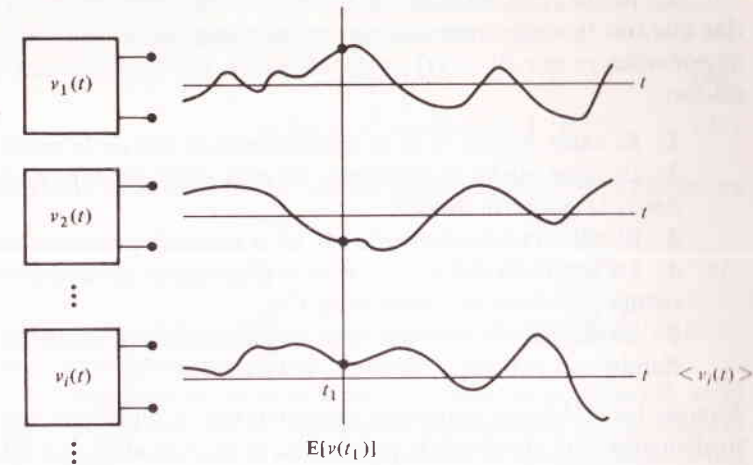



FIGURA 3.7
Promedios de tiempo y promedios de ensamble.

En forma alterna, supóngase que se pudiera disponer de un gran número de generadores idénticos, lo que se conoce como un *ensamble*. Los miembros del ensamble son idénticos en el sentido de que tienen la misma descripción estadística; sin importar que sus señales individuales —conocidas como funciones muestra— sean diferentes. Una señal en particular, designada ahora $v_i(t)$ para claridad, es una de la infinidad de funciones muestra que un determinado generador de ruido puede producir. La familia completa de funciones muestra se simbolizará con $v(t)$.

La figura 3.7 ilustra esta noción e indica que se incluyen *dos* clases diferentes de promedios: se pueden efectuar mediciones sucesivas en una función muestra $v_i(t)$ para encontrar sus *promedios de tiempo* $\langle v_i(t)\rangle$, etc.; o se pueden examinar todas las funciones muestra durante un tiempo particular, digamos t_1, y encontrar los *promedios de ensamble* $E[v(t_1)]$, etc. De manera obvia, los promedios de ensamble son los mismos que los promedios estadísticos en $t = t_1$. Pero, ¿cómo se relacionan con los promedios de tiempo de una función muestra?

En general, los promedios de ensamble pueden diferir de tales promedios de tiempo. Seguramente serían diferentes si las propiedades estadísticas de los generadores estuvieran cambiando con el tiempo, ya que la variación en el tiempo sería promediada en el tiempo pero no en el ensamble. Más aún, los promedios de ensamble a diferentes tiempos, también serían diferentes.

Sin embargo, muchas de las señales aleatorias en los sistemas de comunicación provienen de *procesos ergódicos*, significando esto que los *promedios de tiempo y de ensamble son idénticos*, o sea, $E[v(t)] = \langle v(t)\rangle$, etc. Un proceso ergódico es también *estacionario*, puesto que los promedios de ensamble son independientes del tiempo de observación.

En lo que resta del texto, se tratará casi en forma exclusiva con señales aleatorias que son funciones muestra de procesos ergódicos. Por lo que estas señales son *de potencia*, ya que $E[v^2(t)]$ no depende de t, y se pueden hacer los siguientes enunciados:

1 El valor *medio* $\bar{v}$ es la *componente* de CD de la señal.
2 El valor *medio al cuadrado* $\bar{v}^2$ es la potencia de la componente de CD, es decir, la *potencia* de CD.
3 El *valor cuadrado medio* $\overline{v^2}$ es la *potencia promedio* total.
4 La *variancia* $\sigma v^2 = \overline{v^2} - \bar{v}^2$ es la potencia en la componente variando con el tiempo, es decir, la *potencia de* CA.
5 La *desviación estándar* σ_v es la raíz cuadrática media de la componente variando con el tiempo, es decir, el *valor rms*.

Aunque las relaciones anteriores propician que el ingeniero electricista se sienta en su elemento en el mundo de la estadística, se debe recalcar que sólo se aplican a casos muy especiales de fuentes ergódicas.†

Correlación

Se tienen dos razones para considerar la función de autocorrelación de una señal aleatoria. Primera, a $R_v(\tau)$ le corresponde proporcionar información útil acerca de $v(t)$. Segunda, según, el teorema de Wiener-Kinchine, la descripción del dominio de la frecuencia de una señal aleatoria es su densidad espectral de potencia $G_v(f) = \mathscr{F}[R_v(\tau)]$. Antes de aceptar esto último, véase cómo se puede interpretar $R_v(\tau)$ cuando $v(t)$ es aleatoria.

Primero se define el *producto escalar* de dos señales aleatorias ergódicas como

$$\langle v(t),w(t)\rangle \triangleq \mathbf{E}[v(t)w^*(t)] \tag{1}$$

por consiguiente, la norma al cuadrado es $\|v\|^2 = E[|v(t)|^2]$. De manera similar

$$R_{vw}(\tau) \triangleq \mathbf{E}[v(t)w^*(t-\tau)] \tag{2}$$

lo cual es semejante a la definición de correlación cruzada de la sección 2.6. Pero ergodicidad significa que son intercambiables los promedios de tiempo y de ensamble. Por lo que, considerando real a $v(t)$, su autocorrelación es

$$R_v(\tau) = \mathbf{E}[v(t)v(t-\tau)] = \langle v(t)v(t-\tau)\rangle \tag{3}$$

en la cual la segunda expresión es un promedio de tiempo y no un producto escalar. En lo sucesivo, se tratará pricipalmente con señales aleatorias que sean reales y ergódicas, de manera que una u otra de las expresiones de la ecuación (3) dé la función de autocorrelación. Se debe notar con cuidado que $R_v(\tau)$ es una función *determinística* aun cuando $v(t)$ es aleatoria.

† Véase Lathi (1968, cap. 3) para un análisis del caso no ergódico.

Es obvio que todas las propiedades matemáticas adjudicadas a $R_v(\tau)$ también se aplican a la autocorrelación de señales aleatorias. En particular

$$R_v(-\tau) = R_v(\tau) \tag{4a}$$

$$R_v(0) = \overline{v^2} = \sigma_v^2 + \bar{v}^2 \geq |R_v(\tau)| \tag{4b}$$

de tal manera que $R_v(\tau)$ es una función par con un máximo en $\tau = 0$. Además, es verdad que

$$\lim_{\tau\to\infty} R_v(|\tau|) = \bar{v}^2 \tag{5}$$

advertidos de que el proceso es no periódico. (Si el proceso es periódico, su autocorrelación es también periódica con el mismo periodo.)

En un sentido, la autocorrelación es una medida tanto de la *variación en el tiempo* como de la *dependencia estadística*. Supóngase, por ejemplo, que se considera un valor de τ muy pequeño comparado con los intervalos de tiempo en los cuales $v(t)$ tiene cambios significativos; entonces $v(t-\tau) \approx v(t)$, y $R_v(\tau) \approx \overline{v^2}(t) = \sigma^2 + v^2$. Conforme $|\tau|$ crece, $R_v(\tau)$ generalmente decrecerá al principio (no puede aumentar), y el intervalo de τ alrededor de $\tau = 0$ para el cual $R_v(\tau) - \overline{v^2} \approx \sigma^2$, es una medida de la variación en el tiempo de $v(t)$. Lo pequeño de este intervalo, ocasiona las variaciones más rápidas. Por otra parte, si $|\tau|$ es muy grande, se podría encontrar que $v(t)$ y $v(t-\tau)$ tienen tan poco en común y son *estadísticamente independientes*; entonces $R_v(\tau) = v(t)v(t-\tau) = v^2$. Cuando $|R_v(\tau)| - v^2 = 0$, $v(t-\tau)$ y $v(t)$, son *no correlacionadas*, pero no es necesario que sean estadísticamente independientes —a menos que el proceso sea *gaussiano*.

EJERCICIO 3.9 Demuéstrese que, para cualesquier t_1 y t_2 constantes,

$$\mathbf{E}[v(t-t_1)v(t-t_2)] = R_v(t_1 - t_2) \tag{6}$$

donde $v(t)$ es real y ergódica. (*Sugerencia*: Sustitúyase el promedio de ensamble por un promedio sobre todo el tiempo; nótese que t es una variable fantasma.)

Densidad espectral de potencia

Volviendo al dominio de la frecuencia, la única descripción razonable de una señal aleatoria es su *densidad espectral de potencia* $G_v(f)$. Esto es así porque el único conocimiento previo que se tiene sobre una señal aleatoria es su función de probabilidad y sus promedios de ensamble, estando esto último relacionado en forma directa con la potencia promedio si la fuente es ergódica.

Recurriendo al teorema de Wiener-Kinchine, la densidad espectral de potencia de la señal aleatoria $v(t)$ es

$$G_v(f) = \mathscr{F}[R_v(\tau)] = \int_{-\infty}^{\infty} \mathbf{E}[v(t)v(t-\tau)]e^{-j2\pi f\tau}\,d\tau \tag{7}$$

En vista de las propiedades de $R_v(\tau)$ se tiene inmediatamente a partir de la ecuación (7) que

$$G_v(-f) = G_v(f) \tag{8a}$$

$$\int_{-\infty}^{\infty} G_v(f)\,df = R_v(0) = \overline{v^2} \tag{8b}$$

de tal manera que $G_v(f)$ es real y par y su área total es igual a $\overline{v^2}$

Generalmente, $G_v(f)$ es una función de la frecuencia suave y continua. Sin embargo, si $\overline{v} \neq 0$, la densidad espectral tiene un impulso en $f = 0$ cuyo peso es igual a la potencia de CD $\overline{v}^2$. Para demostrar esto, hágase $w(t) = v(t) - \overline{v}$ tal que $\overline{w} = 0$. Entonces, puesto que $v(t) = w(t) + \overline{v}$,

$$\begin{aligned} R_v(\tau) &= \mathbf{E}[w(t)w(t-\tau) + w(t)\bar{v} + \bar{v}w(t-\tau) + \bar{v}^2] \\ &= R_w(\tau) + 2\bar{w}\bar{v} + \bar{v}^2 = R_w(\tau) + \bar{v}^2 \end{aligned} \tag{9a}$$

y la transformación de Fourier conduce a

$$G_v(f) = G_w(f) + \bar{v}^2\,\delta(f) \tag{9b}$$

El ejemplo siguiente, escogido para mantener la confianza del lector en el teorema de Wiener-Kinchine para señales aleatorias, también conduce a una densidad espectral de potencia impulsiva.

Ejemplo 3.6 Sinusoide en fase en forma aleatoria

Considérese el proceso aleatorio

$$v(t) = A\cos(\omega_0 t + \Theta)$$

donde A y $\omega_0 = 2\pi f_0$ son constantes, mientras que Θ es una variable aleatoria sobre el ensamble con una función de densidad de probabilidad uniforme

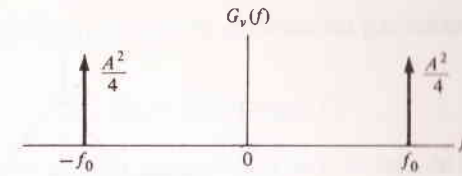
$$p(\theta) = \frac{1}{2\pi} \qquad -\pi \le \theta \le \pi$$

Empleando identidades trigonométricas para desarrollar $v(t)v(t-\tau)$, la función de autocorrelación es

$$\begin{aligned} R_v(\tau) &= \mathbf{E}\left[\frac{A^2}{2}\cos\omega_0\tau + \frac{A^2}{2}\cos(2\omega_0 t - \omega_0\tau + 2\Theta)\right] \\ &= \frac{A^2}{2}\mathbf{E}[\cos\omega_0\tau] + \frac{A^2}{2}\mathbf{E}[\cos(2\omega_0 t - \omega_0\tau + 2\Theta)] \end{aligned}$$

Pero $E[\cos\omega_0\tau] = \cos\omega_0\tau$ puesto que esta magnitud no es aleatoria, en tanto que

$$\mathbf{E}[\cos(2\omega_0 t - \omega_0\tau + 2\Theta)] = \int_{-\infty}^{\infty}\cos(2\omega_0 t - \omega_0\tau + 2\theta)p(\theta)\,d\theta$$

FIGURA 3.8
Densidad espectral de potencia de una sinusoide con fase aleatoria.

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}\cos(2\omega_0 t - \omega_0\tau + 2\theta)\,d\theta = 0$$

Por lo tanto

$$R_v(\tau) = \frac{A^2}{2}\cos\omega_0\tau \tag{10a}$$

la cual es idéntica a la autocorrelación de una sinusoide *determinística* —véase el ejemplo 2.15, sección 2.6.

La transformación de Fourier da entonces una densidad espectral de potencia impulsiva

$$G_v(f) = \frac{A^2}{4}\delta(f - f_0) + \frac{A^2}{4}\delta(f + f_0) \tag{10b}$$

como se indica en la figura 3.8. Se ve que esta señal aleatoria tiene cantidades específicas de potencia promedio concentradas en $f = \pm f_0$ y no habrá densidad espectral en ninguna otra parte, de acuerdo con el razonamiento inicial. Además, $\int_{-\infty}^{\infty} G_v(f)\,df = A^2/2$ lo cual es claro que es igual a $\overline{v^2}$, y el hecho de que no hay impulso en $f = 0$ concuerda con $\overline{v} = 0$.

Ejemplo 3.7 Onda binaria aleatoria ★

El ejemplo anterior es bastante engañoso debido a su sencillez. Muy a menudo determinar una función de autocorrelación es un problema analítico tenaz —como lo demostrará el ejemplo.

La figura 3.9 es una función muestra de la onda binaria aleatoria $v(t)$. Durante cualquier intervalo $(n-1)T < t - T_d < nT$, $v(t)$ adquiere uno u otro de los va-

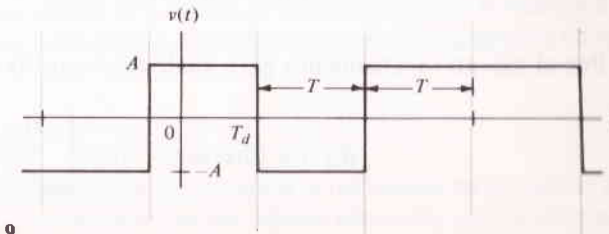



FIGURA 3.9
Función muestra de una onda binaria aleatoria.

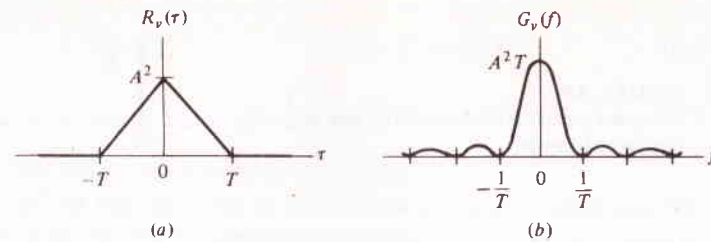



FIGURA 3.10
Onda binaria aleatoria, *a*) Autocorrelación; *b*) densidad espectral de potencia.

lores $+A$ o $-A$. Estas dos amplitudes tienen igual probabilidad, y la amplitud en uno cualquiera de los intervalos es independiente de todos los otros. El término de retardo T_d es una variable aleatoria sobre el ensamble, distribuido de manera uniforme sobre $[0, T]$.

Para encontrar la autocorrelación, primero considérese $|\tau| > T$; así, para cualquier función muestra, $v(t)$ y $v(t-\tau)$ están en intervalos diferentes y, por lo tanto, son *independientes*. Así

$$\mathbf{E}[v(t)v(t-\tau)] = \bar{v}^2 = \langle v(t)\rangle^2 = 0$$

donde el promedio del tiempo $\langle v(t)\rangle = 0$ se obtiene por inspección de $v(t)$.

Ahora, considérese $|\tau| < T$ y hágase $t = 0$; entonces $v(0)$ y $v(-\tau)$ están en el mismo intervalo sólo si $T_d - T < -|\tau|$. Así

$$\mathbf{E}[v(0)v(-\tau)] = \begin{cases} A^2 & T_d < T - |\tau| \\ 0 & \text{de otra manera} \end{cases}$$

La probabilidad de que $T_d < T - |\tau|$ es

$$P(T_d < T - |\tau|) = \int_0^{T-|\tau|} \frac{dt_d}{T} = \frac{T - |\tau|}{T}$$

de tal manera que

$$\mathbf{E}[v(0)v(-\tau)] = A^2 P(T_d < T - |\tau|) = A^2\left(1 - \frac{|\tau|}{T}\right)$$

Por el mismo razonamiento para cualquier valor de t

$$R_v(\tau) = \mathbf{E}[v(t)v(t-\tau)] = \begin{cases} A^2\left(1 - \dfrac{|\tau|}{T}\right) & |\tau| < T \\ 0 & |\tau| > T \end{cases}$$

$$= A^2\Lambda\left(\frac{\tau}{T}\right)$$

donde $\Lambda(\tau/t)$ es la función triángulo. Por último, se obtiene el conocido par transformado

$$G_v(f) = A^2 T \operatorname{senc}^2 fT$$

La figura 3.10 muestra la función de autocorrelación y el espectro de potencia de la onda binaria aleatoria. ////

Señales aleatorias moduladas

Cuando se analizan sistemas de comunicación, se encuentran a menudo procesos modulados de la forma $v(t)\cos\omega_0 t$, donde $v(t)$ es una señal aleatoria. Aun cuando $v(t)$ es ergódico, se demuestra con facilidad que el proceso modulado es no ergódico. Sin embargo, se puede tratar con el ensamble

$$z(t) = v(t)\cos(\omega_0 t + \Theta) \qquad (11a)$$

lo cual será ergódico si Θ es una variada independiente distribuida de manera uniforme sobre $[-\pi, \pi]$. Agregando Θ a esta modalidad, simplemente se reconoce la selección arbitraria del origen del tiempo cuando $v(t)$ y $\cos\omega_0 t$ provienen de fuentes físicamente independientes, como es siempre el caso.

Para encontrar $R_z(\tau)$ en términos de $R_v(\tau)$, se siguen de manera esencial los mismos lineamientos que en el ejemplo 3.6 excepto en el sentido de que los promedios de ensamble deben efectuarse tanto con respecto a $v(t)$ como con respecto a Θ. Por lo tanto, desarrollando $z(t)z(t-\tau)$,

$$\begin{aligned} R_z(\tau) &= \tfrac{1}{2}\mathbf{E}[v(t)v(t-\tau)\cos\omega_0\tau + v(t)v(t-\tau)\cos(2\omega_0 t - \omega_0\tau + 2\Theta)] \\ &= \tfrac{1}{2}\mathbf{E}[v(t)v(t-\tau)]\{\mathbf{E}[\cos\omega_0\tau] + \mathbf{E}[\cos(2\omega_0 t - \omega_0\tau + 2\Theta)]\} \\ &= \tfrac{1}{2}R_v(\tau)\cos\omega_0\tau \end{aligned} \qquad (11b)$$

ya que $E[\cos\omega_0\tau] = \cos\omega_0\tau$ en tanto que $E[\cos(2\omega_0 t - \omega_0\tau + 2\Theta)] = 0$, como se determinó con anterioridad. Haciendo $\tau = 0$ se llega a $R_z(0) = \tfrac{1}{2}R_v(0)$, lo cual es consistente con el hecho de que $E[z^2(t)] = \tfrac{1}{2}E[v^2(t)]$.

Aplicando el teorema de Wiener-Kinchine, $G_z(f) = \mathscr{F}[R_v(\tau)]$, se tiene

$$G_z(f) = \tfrac{1}{4}[G_v(f - f_0) + G_v(f + f_0)] \qquad (11c)$$

por lo que la modulación ha trasladado a la densidad espectral $G_v(f)$ por encima y debajo en f_0 unidades; un caso similar al teorema de la modulación de la sección 2.3. Nótese también las similitudes y diferencias entre este caso y el de la sinusoide en fase de manera aleatoria.

Señales aleatorias incoherentes

Los conceptos de la ortogonalidad y de la incoherencia se definieron en la sección 2.6, para señales determinísticas. Aquí se dice que dos señales aleatorias son *incoherentes* si son *ortogonales para toda* τ, es decir,

$$R_{vw}(\tau) = R_{wv}(\tau) = 0 \qquad (12)$$

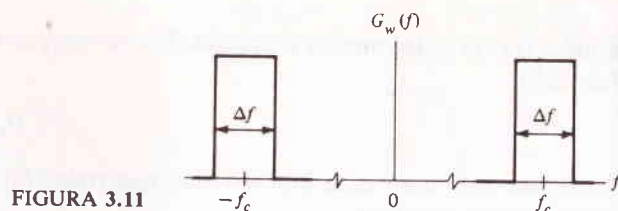



FIGURA 3.11

En relación con la ecuación (1), una condición suficiente para la incoherencia es que $v(t)$ y $w(t)$ sean *estadísticamente independientes* y una o la otra tengan *valor medio cero*, es decir, $\overline{wv} = 0$. Cuando se aplica la ecuación (12), la señal $z(t) = v(t) + w(t)$ tiene

$$R_z(\tau) = R_v(\tau) + R_w(\tau) \qquad (13a)$$

y haciendo $\tau = 0$,

$$\overline{z^2} = \overline{v^2} + \overline{w^2} \qquad (13b)$$

Más aún, a partir de la linealidad de la transformación de Fourier,

$$G_z(f) = G_v(f) + G_w(f) \qquad (13c)$$

por lo que se tiene superposición de densidad espectral de potencia, así como superposición de potencia promedio.

Estos resultados serán valiosos de manera especial cuando se estudie el ruido y señales contaminadas con ruido.

Representación en serie de Fourier de una señal aleatoria

Las funciones de autocorrelación y de densidad espectral, son representaciones determinísticas de una señal aleatoria en el dominio de τ y en el de f, respectivamente. Hay también varias representaciones en series del dominio del tiempo, siendo la más importante la serie de Fourier. Sin embargo, tales representaciones son aproximadas e incluyen coeficientes no determinísticos (es decir, aleatorios).

Considérese para comenzar, la señal aleatoria $w(t)$ que tiene $G_w(f)$ como se muestra en la figura 3.11, donde $\Delta f \ll f_c$. Su función de autocorrelación es

$$R_w(\tau) = \mathscr{F}^{-1}[G_w(f)]$$
$$= 2G_w(f_c)\Delta f \operatorname{senc} \tau\Delta f \cos \omega_c \tau$$

puesto que $G_w(f)$ consiste de dos funciones rectangulares de la misma altura, $G_w(f_c) = G_w(-f_c)$. Ahora, si se concentra la atención a valores de τ relativamente pequeños, o sea $|\tau| \ll 1/\Delta f$, entonces senc $\tau\Delta f \approx 1$ y

$$R_w(\tau) \approx [2G_w(f_c)\Delta f] \cos \omega_c \tau \qquad |\tau| \ll \frac{1}{\Delta f} \qquad (14a)$$

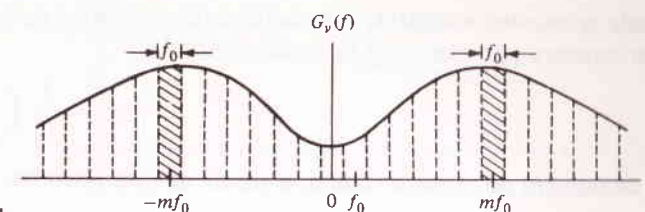



FIGURA 3.12
Espectro de potencia fraccionado.

Comparando esto con la ecuación (10*a*) del ejemplo 3.6, sugiere la aproximación

$$w(t) \approx A \cos (\omega_c t + \theta) \qquad (14b)$$

donde A y Θ son variables aleatorias. Haciendo $\tau = 0$ en la ecuación (14*a*) se demuestra que $\overline{w^2} = 2G_w(f_c)\Delta f$, por lo que

$$\overline{A^2} = 4G_w(f_c)\Delta f \qquad (14c)$$

Supóngase todavía que Θ está distribuida de manera uniforme, puesto que $G_w(f)$ no proporciona información de fase. En esta forma, se ha aproximado $w(t)$ a una sinusoide a la frecuencia f_c con amplitud y fase aleatorias; por lo pequeño del valor de Δf, se tiene la mejor aproximación. En realidad, si $\Delta f \to 0$, $w(t)$ debe ser una sinusoide, pero con amplitud sumamente pequeña.

La aproximación anterior se puede extender a una $G_v(f)$ arbitraria fraccionándola en bandas muy angostas de ancho $\Delta f = f_0$ con centro en $f = mf_0$, $m = 0, \pm 1, \pm 2, \ldots$, como se muestra en la figura 3.12. Cada *par* de bandas en $f = \pm mf_0$ corresponde a

$$v_m(t) \approx A_m \cos (m\omega_0 t + \theta_m) \qquad (15a)$$

donde Θm está distribuido de manera uniforme y, de la ecuación (14*c*),

$$\overline{A_m^{\,2}} = 4f_0 \, G_v(mf_0) \qquad (15b)$$

Entonces, dado que no se traslapan dos pares diferentes de bandas, se puede escribir

$$R_v(\tau) = \mathscr{F}^{-1}[G_v(f)]$$
$$= 2G_v(f_c)\Delta f \operatorname{senc} \tau\Delta f \cos \omega_c \tau \qquad (16)$$

y todos los términos de la suma son incoherentes entre sí.

La ecuación (16) es un desarrollo en serie de Fourier con periodo $T_0 = 1/f_0$ puesto que todas las componentes son armónicas de f_0. Así, esta aproximación es válida en el intervalo T_0 segundos de duración, un intervalo que se puede hacer arbitrariamente grande haciendo a f_0 pequeño también en forma arbitraria, lo cual mejora también la aproximación.

Convirtiendo la ecuación (16) a forma exponencial se llega a

$$v(t) \approx \sum_{m=-\infty}^{\infty} c_m e^{jm\omega_0 t} \qquad (17a)$$

donde se pueden encontrar los coeficientes aleatorios c_m a partir de cualquier función muestra por medio de la regla usual

$$c_m = \frac{1}{T_0} \int_{T_0} v(t) e^{-jm\omega_0 t}\, dt \qquad (17b)$$

Se demuestra de manera rápida, a partir de esta ecuación que

$$\overline{c_m} = \begin{cases} \bar{v} & m = 0 \\ 0 & m \neq 0 \end{cases} \qquad (18a)$$

pero las otras propiedades de c_m se encuentran de manera más fácil al notar que

$$A_m = 2|c_m| \qquad \theta_m = \arg\,[c_m]$$

Por lo que†

$$\mathrm{E}[|c_m|^2] = f_0\, G_v(mf_0) \qquad (18b)$$

$$\mathrm{E}[c_m c_n] = 0 \qquad m \neq n \qquad (18c)$$

En resumen, se ha desarrollado una representación en serie aproximada para $v(t)$ durante un intervalo $T_0 = 1/f_0$. Los coeficientes de la serie son variables aleatorias no correlacionadas, cuyos promedios estadísticos están relacionados con la densidad espectral.

3.6 RUIDO Y SU FILTRACIÓN

Las señales eléctricas no deseadas provienen de una diversidad de fuentes, clasificadas de manera general como interferencia hecha por el hombre o ruido que ocurre en forma natural. La interferencia hecha por el hombre proviene de otros sistemas de comunicación, como chispas de ignición en los automóviles o en los conmutadores, zumbido de 60 Hz, etc.; los fenómenos generadores de ruido natural comprenden disturbios atmosféricos, radiación extraterrestre y ruido en circuitos. Los efectos de muchas señales no deseadas, se pueden reducir o eliminar en forma completa, por medio de un cuidadoso diseño de ingeniería. Pero siempre persistirán ciertas señales aleatorias inevitables, las cuales presentan un límite fundamental a la elaboración de los sistemas.

Una causa inevitable de ruido eléctrico es el movimiento térmico de los electrones en los medios de conducción (alambres, resistencias, etc.). Mientras los sistemas de comunicación sean construidos de tal material, persistirá este *ruido térmico*. Por consiguiente, esta sección da principio con una breve discusión sobre el ruido térmico, el cual a su vez, conduce a la abstracción que es el ruido blanco. Se considera después el ruido blanco filtrado y las relaciones de entrada y salida. En subsecuentes capítulos y en el Apéndice B se desarrollan otros aspectos de análisis de ruido.

† En realidad, las ecuaciones (18*b*) y (18*c*) son exactas sólo en el límite cuando $T_0 \to \infty$. Véase Thomas (1969, cap. 4).

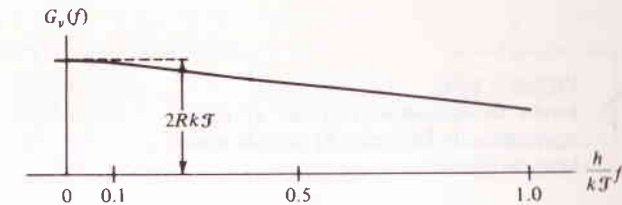



FIGURA 3.13
Densidad espectral de ruido térmico, V²/Hz.

Ruido térmico

El ruido térmico, según los propósitos de este texto, es el voltaje de ruido debido al movimiento aleatorio de partículas cargadas (por lo general electrones) en medios conductores. Según la teoría cinética, la energía promedio de una partícula a la temperatura absoluta $\mathcal{T}$ es proporcional a $k\mathcal{T}$, siendo k la constante de Boltzmann. Se esperan así, valores de ruido térmico que incluyan el producto $k\mathcal{T}$; en efecto, se obtendrá una medida de potencia de ruido en términos de temperatura.

Johnson (1928) y Nyquist (1928) hicieron historia al ser los primeros en estudiar el ruido térmico en resistencias metálicas, de lo cual proviene la designación de *ruido de Johnson o ruido de resistencia*. Sin embargo, en cualesquier medios pasivos lineales, se encuentra el mismo fenómeno básico. Se omite aquí la deducción general de las propiedades del ruido térmico, pero, gracias a estudios teóricos y experimentales intensivos, existe un amplio conocimiento sobre el tema, el que se explotará con libertad.

Cuando una resistencia metálica R se encuentra a la temperatura $\mathcal{T}$, existe un voltaje aleatorio $v(t)$ producido en las terminales a circuito abierto a causa del movimiento electrónico aleatorio. Consistente con el teorema del límite central, $v(t)$ tiene una *distribución gaussiana* con $\bar{v} = 0$ y

$$\overline{v^2} = \sigma_v^2 = \frac{2(\pi k\mathcal{T})^2}{3h} R \qquad \text{volts al cuadrado } (V^2) \qquad (1)$$

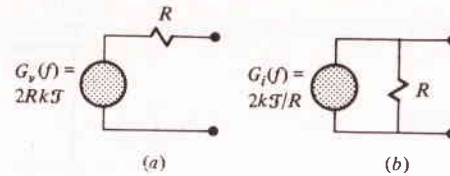
donde

k = constante de Boltzmann = 1.37×10^{-23} joules por grado (J/grado)
h = constante de Planck = 6.62×10^{-34} joule segundo (J · s)

y $\mathcal{T}$ se mide en Kelvin. La presencia de la constante de Planck en la ecuación (1) indica un resultado de mecánica cuántica. Yendo más lejos, se demuestra en mecánica cuántica que la densidad espectral del ruido térmico es

$$G_v(f) = \frac{2Rh|f|}{e^{h|f|/k\mathcal{T}} - 1} \qquad \text{volts al cuadrado por hertz } (V^2/\text{Hz}) \qquad (2)$$

como se muestra en la gráfica de la figura 3.13 para $f \geq 0$.

FIGURA 3.14
Ruido de resistencia térmica. *a*) circuito equivalente de Thévenin; *b*) circuito equivalente de Norton.

En realidad, los ingenieros de comunicaciones casi nunca han tenido necesidad de estas dos expresiones. Para ver la razón de esto, se considera la temperatura de salón o *temperatura normal* como

$$\mathcal{T}_0 \triangleq 290°\mathrm{K}\ (63°\mathrm{F}) \qquad (3a)$$

lo cual está más en el lado frío pero simplifica el trabajo numérico, puesto que

$$k\mathcal{T}_0 \approx 4 \times 10^{-21} \quad \text{watts segundo (W} \cdot \text{s)} \qquad (3b)$$

Entonces, si la resistencia está en $\mathcal{T}_0$, $G_v(f)$ es *constante* de manera esencial para $|f| < 0.1\, k\mathcal{T}_0/h \approx 10^{12}$ Hz. Pero este límite superior cae en la porción infrarroja del espectro electromagnético, más arriba del punto donde los componentes eléctricos convencionales cesan de responder. Aun a temperaturas criogénicas ($\mathcal{T} \approx 0.01\mathcal{T}_0$) vale esta conclusión.

Por lo tanto, a todos los intentos y propósitos, se puede decir que la densidad espectral del ruido térmico es constante, o sea,

$$G_v(f) = 2Rk\mathcal{T} \qquad \mathrm{V}^2/\mathrm{Hz} \qquad (4)$$

como se obtiene a partir de la ecuación (2) con $h|f|/k\mathcal{T} \ll 1$. La única dificultad con la ecuación (4) es que ésta predice en forma errónea que $\overline{v^2} = \infty$ cuando $G_v(f)$ se integra sobre toda la f; se debe llevar en mente, sin embargo, que nunca se tratará en forma directa con $\overline{v^2}$ puesto que $v(t)$ siempre estará sujeta a los efectos de filtrado de otros circuitos. Este tema se examinará pronto. Sin embargo, primero se empleará la ecuación (4) para construir el modelo equivalente de Thévenin de una resistencia térmica, figura 3.14*a*. Aquí la resistencia se reemplaza por una resistencia *sin ruido* ficticia del mismo valor, y el ruido se representa por medio de un generador de voltaje de valor cuadrático medio. De manera similar, la figura 3.14*b* es el equivalente de Norton con un generador de corriente de valor cuadrático medio que tiene $G_i(f) = G_v(f)/R^2 = 2k\mathcal{T}/R$. Ambos generadores están sombreados para indicar su naturaleza especial.

En lugar de tratar con voltajes o corrientes cuadráticos medios, para definir el ruido térmico por medio de su *potencia disponible* se depura la notación y se agiliza el cálculo. La potencia disponible es la potencia *máxima* que se puede entregar a una carga de una fuente que tiene resistencia fija diferente de cero. El ya conocido teorema de la transferencia de potencia máxima establece que esta potencia se entrega sólo cuando la impedancia de carga es el complejo conjugado de la impedancia de la fuente. Se dice entonces que la carga está *acoplada* a la fuente, una condición generalmente buscada en un sistema de comunicación.

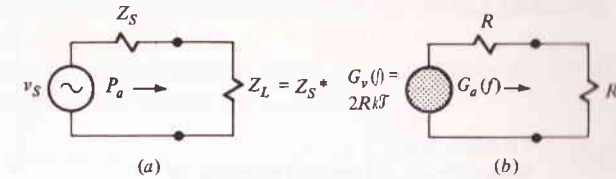



FIGURA 3.15
Potencia disponible. *a*) Fuente de señal con carga acoplada; *b*) resistencia térmica con carga acoplada.

Sea una fuente de señal convencional con una impedancia $Z_s = R_s + jX_s$, y sea el voltaje a circuito abierto v_s, figura 3.15*a*. Si la carga está acoplada de tal manera que $Z_L = Z_s^* = R_s - jX_s$, entonces el voltaje terminal es $v_s/2$, y la potencia disponible es

$$P_a = \frac{\langle [v_S(t)/2]^2 \rangle}{R_S} = \frac{\langle v_S^2(t) \rangle}{4R_S}$$

Empleando el modelo de Thévenin, se puede extender este concepto a un resistor térmico visto como una fuente de ruido, figura 3.15*b*. Por comparación, la *densidad espectral disponible* en la resistencia de carga es

$$G_a(f) = \frac{G_v(f)}{4R} = \tfrac{1}{2}k\mathcal{T} \qquad \text{watts por hertz (W/Hz)} \qquad (5)$$

lo cual depende sólo de la temperatura. Por lo tanto, ¡un resistor térmico entrega un máximo de $k\mathcal{T}/2$ W/Hz a una carga acoplada, sin importar el valor de R!

Ruido blanco y temperatura de ruido

Además de los resistores térmicos, muchos otros tipos de fuentes de ruido son gaussianos y tienen una densidad espectral plana sobre un intervalo amplio de frecuencias. Tal espectro tiene todos los componentes de frecuencia en igual proporción y se le designa en forma correcta como ruido *blanco*, por analogía con la luz blanca.

La densidad espectral del ruido blanco se escribe en general como,

$$G(f) = \frac{\eta}{2} \qquad (6a)$$

donde el factor en apariencia extraño ½ se incluye para indicar que se asocia la mitad de la potencia con la frecuencia positiva y la mitad con la frecuencia negativa (fi-

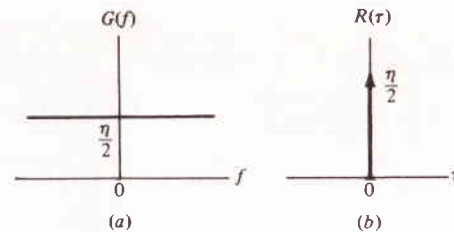



FIGURA 3.16
Ruido blanco. *a*) Densidad espectral de potencia; *b*) autocorrelación.

gura 3.16*a*). En forma alterna η es la densidad de potencia de *frecuencia positiva*. Puesto que se conoce $G(f)$, la función de autocorrelación se obtiene inmediatamente por la transformación de Fourier

$$R(\tau) = \int_{-\infty}^{\infty} \frac{\eta}{2} e^{j\omega\tau}\, df = \frac{\eta}{2}\,\delta(\tau) \qquad (6b)$$

como en la figura 3.16*b*.

A partir de la ecuación (6*b*) o de la figura 3.16*b*, se observa que $R(\tau \neq 0) = 0$, de tal manera que dos muestras direrentes cualesquiera de una señal de ruido blanco gaussiano son *no correlacionadas* y, por lo tanto, *estadísticamente independientes*. Esta observación, con el espectro de potencia constante, conduce a una conclusión interesante: si el ruido blanco se expone en un osciloscopio, los barridos sucesivos son siempre diferentes de los otros; sin embargo, la forma de onda siempre tiene la misma "apariencia", sin importar la velocidad de barrido que se use, puesto que todos los grados de variación de tiempo (componentes de frecuencia) están contenidos en igual proporción. De manera similar, si el ruido blanco impulsa un altoparlante, éste siempre sonará igual, algo así como una caída de agua. Conforme lo expone Pierce†

> En forma matemática, el ruido blanco gaussiano . . . es el resumen de los diferentes e inesperados. Es el menos predecible y el más original de los sonidos. Sin embargo, a un ser humano todos los sonidos de ruido blanco gaussiano le son iguales. Sus sutilezas se le esconden, y dice de él que es pesado y monótono.

El valor de η en las ecuaciones (6*a*) y (6*b*) depende de dos factores: el tipo de la fuente del ruido y el tipo de densidad espectral, es decir, el voltaje cuadrático medio, la corriente cuadrática media o la potencia disponible. Si la fuente es un *resistor térmico*, entonces

$$\eta_v = 4Rk\mathcal{T} \qquad \eta_i = \frac{4k\mathcal{T}}{R} \qquad \eta_a = k\mathcal{T} \qquad (7)$$

donde el tipo de densidad espectral se indica por medio de subíndices. Además, cualquier fuente de ruido *término* tiene, por definición, $\eta_a = k\mathcal{T}$, Otras fuentes de ruido

† J. R. Pierce, *Symbols, Signals, and Noise*, Harper & Row, Publishers, Incorporated, Nueva York, 1961.

blanco son *no térmicas* en el sentido de que la potencia disponible no está relacionada con la temperatura física.
No obstante, se puede hablar de la *temperatura de ruido* F_N de cualquier fuente de ruido blanco, térmico o no térmico, por medio de la definición.

$$\mathcal{T}_N \triangleq \frac{2G_a(f)}{k} = \frac{\eta_a}{k} \qquad (8a)$$

donde $\eta_a/2$ es la máxima potencia de ruido que la fuente puede entregar por unidad de frecuencia. Entonces, dada una temperatura de ruido de fuente,

$$\eta_a = k\mathcal{T}_N \qquad (8b)$$

Se debe recalcar que $\mathcal{T}_N$ no es necesariamente una temperatura física. Por ejemplo, ciertos generadores de ruido electrónicos tiene $\mathcal{T}_N \approx 10\mathcal{T}_0 \approx$ 3 000 K(5 000° F), pero es obvio que los dispositivos no pueden estar a esa temperatura.

Ruido blanco filtrado

El modelo de ruido blanco resulta razonable siempre y cuando se relacione con la salida de un filtro y la densidad espectral sea más o menos constante sobre la banda de paso, una situación muy común en sistemas de comunicación. Para el estudio del ruido blanco filtrado —o filtración de señales aleatorias en general— la relación de espectros de potencia de entrada y salida

$$G_y(f) = |H(f)|^2 G_x(f) \qquad (9a)$$

es una herramienta básica, donde $H(f)$ es la función de transferencia del filtro. Así, si la entrada a un sistema lineal invariante en el tiempo, es la señal aleatoria $x(t)$, la salida será una señal aleatoria $y(t)$ con $G_y(f)$ como en la ecuación (9*a*), y

$$R_y(\tau) = \int_{-\infty}^{\infty} |H(f)|^2 G_x(f) e^{j\omega\tau}\, df \qquad (9b)$$

$$\overline{y^2} = R_y(0) = \int_{-\infty}^{\infty} |H(f)|^2 G_x(f)\, df \qquad (9c)$$

Se supone, por consiguiente, que $x(t)$ proviene de un proceso ergódico, en cuyo caso también $y(t)$ será ergódico.

Para ilustrar; si el ruido blanco es la entrada de un filtro pasabajas ideal de ganancia unitaria y de ancho de banda B, entonces

$$G_y(f) = \frac{\eta}{2}\,\Pi\left(\frac{f}{2B}\right) \qquad (10a)$$

El espectro de potencia de salida es, de esta manera, una función rectangular (figura 3.17*a*. Además, la autocorrelación de la salida es una función senc (figura 3.17*b*); de manera especifica

$$R_y(\tau) = \eta B \operatorname{senc} 2B\tau \qquad (10b)$$

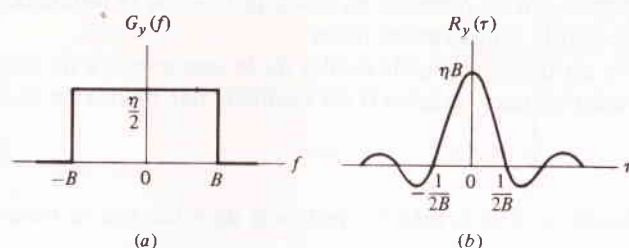



FIGURA 3.17
Ruido blanco pasado por un filtro pasabajas ideal. *a*) Espectro de potencia; *b*) autocorrelación.

Como se evidencia en la figura, el proceso de filtrado ha dado lugar a tres cosas:

1 El espectro de potencia no es mayor que el blanco, aunque es constante sobre un intervalo finito de frecuencia.
2 La potencia de salida es finita. En efecto, $\overline{y^2} = \eta B$.
3 La señal de potencia de salida está correlacionada a intervalos aproximados. $1/2B$.

Aunque las conclusiones anteriores tuvieron como base los filtros pasabajos ideales, se obtienen resultados similares con cualquier filtro real. Se debe notar en forma particular, que el espectro del ruido blanco filtrado se mezcla con el *contorno* de $|H(f)|^2$. Puesto que el espectro resultante no es mayor que el blanco, al ruido blanco filtrado se le designa a menudo como *ruido coloreado*.

También sería deseable conocer la descripción probabilística de la señal filtrada, digamos su función de densidad de probabilidad, y de esta observación obtener tanto buenas como malas noticias. Las malas noticias son éstas: no hay reglas generales que relacionen a las funciones de densidad de probabilidad de entrada y salida, salvo para un caso especial. En este caso especial están las buenas noticias, a saber:

> Si la entrada a un sistema lineal invariante en el tiempo, es una señal aleatoria *gaussiana*, entonces la salida es también *gaussiana*.

Esto proviene del hecho de que cualquier transformación lineal de una variada gaussiana conduce a otra variada gaussiana; pueden cambiarse los promedios estadísticos, pero no el modelo de probabilidad. Se considera este caso especial como una buena noticia, simplemente porque el modelo gaussiano es válido para muchas (pero no todas) señales aleatorias que hay en la ingeniería de la comunicación.

Ejemplo 3.8 Ruido térmico en un circuito RC

Para conllevar varios de los temas tratados hasta aquí, considérese el circuito RC de la figura 3.18*a* donde el resistor está a la temperatura $\mathcal{T}$. Sustituyendo este resistor por su modelo de Thévenin, se llega a la figura 3.18*b*, que es una fuente de ruido blanco con $G_x(f) = 2Rk\mathcal{T}\,\mathrm{V^2/Hz}$

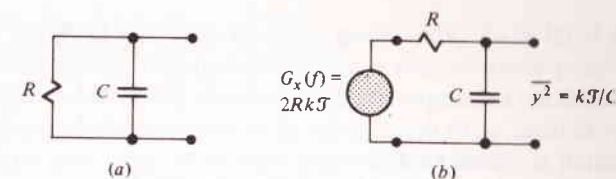



FIGURA 3.18
Un circuito RC con ruido de resistencia. *a*) Diagrama de circuito; *b*) circuito equivalente de ruido.

aplicada a un filtro pasabajos RC *sin ruido*. Puesto que $|H(f)|^2 = [1 + (f/B)^2]^{-1}$, la densidad espectral de salida es

$$G_y(f) = |H(f)|^2 G_x(f) = \frac{2Rk\mathcal{T}}{1 + (f/B)^2} \qquad B = \frac{1}{2\pi RC} \tag{11a}$$

y la autocorrelación de $y(t)$ es

$$R_y(\tau) = 2Rk\mathcal{T}\pi B e^{-2\pi B|\tau|} = \frac{k\mathcal{T}}{C} e^{-|\tau|/RC} \tag{11b}$$

como se obtuvo de un par transformado en la Tabla **A**. $G_y(f)$ y $R_y(\tau)$ se ven en la gráfica de la figura 3.19, la última de las cuales muestra que el intervalo sobre el que el ruido filtrado tiene una correlación apreciable es aproximadamente igual a la constante RC de tiempo de circuito, como se podría sospechar.

Se puede decir además que $y(t)$ es una señal aleatoria gaussiana con $\overline{y} = 0$ (es decir, sin componente de CD) —puesto que $x(t)$ es gaussiana con valor medio cero— y

$$\overline{y^2} = R_y(0) = \frac{k\mathcal{T}}{C} \tag{12}$$

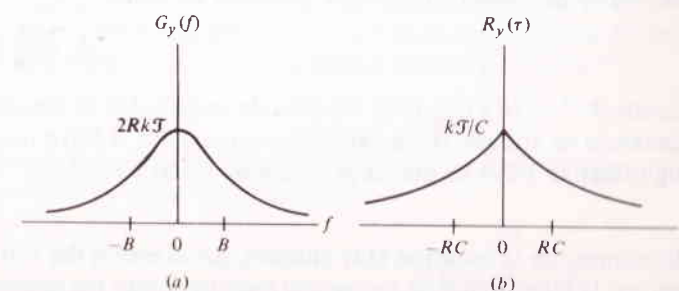
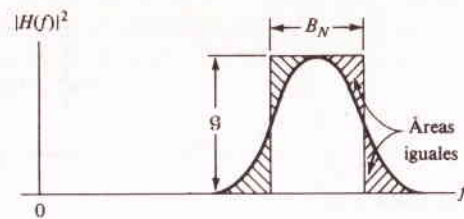



FIGURA 3.19

FIGURA 3.20
Ancho de banda equivalente de ruido B_N de un filtro pasabanda.

De manera soprendente, $\overline{y^2}$ ¡depende de C pero no de R, aunque la fuente de ruido es el resistor térmico! Este resultado en apariencia extraño será explicado adelante; aquí se concluye este ejemplo con un cálculo numérico.

Supóngase que el resistor está a la temperatura ambiente $\mathcal{T}_0$ y $C = 0.1$ μF; entonces

$$\overline{y^2} = \frac{4 \times 10^{-21}}{10^{-7}} = 4 \times 10^{-14} \qquad \text{V}^2$$

y el voltaje de salida rms es $\sigma_y = 2 \times 10^{-7} = 0.2$ microvolts (μV). Estos valores excesivamente pequeños son característicos en ruido térmico, lo cual explica por qué no se nota en situaciones ordinarias. Sin embargo, la señal recibida en un sistema de comunicación de larga distancia puede ser del mismo orden de magnitud o incluso menor, lo cual muestra por qué un ruido térmico es una limitación fundamental en la comunicación eléctrica.

Ancho de banda equivalente de ruido

El ruido blanco filtrado por lo general tiene potencia finita. Para recalcar esto, se designa a la *potencia de ruido promedio* con $N = \overline{y^2}$ y se escribe la ecuación (9*c*) en la forma

$$N = \int_{-\infty}^{\infty} |H(f)|^2 \frac{\eta}{2} df = \eta \int_0^{\infty} |H(f)|^2 \, df$$

Haciéndose notar que la integral depende sólo de la función de transferencia del filtro, por lo que se puede simplificar el análisis de la potencia del ruido definiendo un *ancho de banda equivalente de ruido* B_N como

$$B_N \triangleq \frac{1}{\mathcal{G}} \int_0^{\infty} |H(f)|^2 \, df \qquad (13)$$

donde $\mathcal{G}^{1/2} = |H(f)|_{\text{máx}}$ es la relación de amplitudes de frecuencia central, es decir, la ganancia de voltaje. (Esta definición supone que el filtro tiene una frecuencia central significativa.) Por lo que la potencia de ruido filtrado es

$$N = \mathcal{G}\eta B_N \qquad (14)$$

El examen de la ecuación (14) muestra que el efecto del filtro se ha separado en dos partes: la *selectividad de frecuencia relativa*, como fue descrita por medio de B_N, y la *ganancia de potencia* (o atenuación), representada por medio de $\mathcal{G}$.

Así, como se ilustra en la figura 3.20 para un filtro pasabanda, B_N es igual al ancho de banda de un filtro rectangular ideal que permitiera el paso de tanta potencia de ruido blanco como el filtro en cuestión, siendo iguales sus ganancias máximas.

Por definición, el ancho de banda equivalente de ruido de un filtro ideal es su ancho de banda real. Para filtros prácticos, B_N es algo mayor que el ancho de banda a 3 dB; por ejemplo, un filtro pasabajos RC tiene $B_N = \pi B2$, donde B_N es aproximadamente 50% mayor que B. Sin embargo, como el filtro resulta más selectivo (con características de corte agudo), su ancho de banda de ruido se aproxima al ancho de banda a 3 dB, y para muchas aplicaciones no se está muy lejos de la realidad al considerarlos iguales.

Resumiendo, si $y(t)$ es un ruido blanco filtrado de valor medio cero, entonces

$$\bar{y} = 0$$

$$\overline{y^2} = \sigma_y{}^2 = N = \mathcal{G}\eta B_N \qquad (15)$$

$$\sigma_y = \sqrt{N} = \sqrt{\mathcal{G}\eta B_N}$$

Esto significa que dada una fuente de ruido blanco, un medidor de potencia promedio (o un medidor de voltaje cuadrático medio) acusará una lectura $\overline{y^2} = N = \eta B_N$, donde B_N es el ancho de banda equivalente de ruido del medidor mismo. Despejando, la densidad de potencia de la fuente se puede deducir por medio de $\eta = N/B_N$, siempre y cuando se esté seguro que el ruido es en realidad blanco dentro del intervalo de respuesta de frecuencia del medidor.

Ejemplo 3.9 Ancho de banda equivalente de ruido de un filtro pasabajas RC

Volviendo al filtro pasabajas RC del ejemplo anterior, la frecuencia central del filtro es $f = 0$, así

$$\mathcal{G} = |H(0)|^2 = 1$$

y

$$B_N = \int_0^{\infty} \frac{df}{1 + (f/B)^2} = \frac{\pi}{2} B = \frac{1}{4RC} \qquad (16)$$

La razón de por qué $\overline{y^2}$ en la ecuación (12) es independiente de R resulta ahora aparente si se escribe $\overline{y^2} = \eta B_N(4Rk\mathcal{T}) \times (1/4 \text{ RC})$. Así, cuando aumenta R aumenta la densidad de ruido η (como debe ser) pero disminuye el ancho de banda del ruido B_N. Estos dos efectos se cancelan uno al otro en forma precisa y $\overline{y^2} = k\mathcal{T}/C$.

EJERCICIO 3.10 Empleando la ecuación (21), sección 2.5, demuéstrese que un filtro pasabajas Butterworth de orden M-ésimo tiene

$$B_N = \frac{\pi B}{2M \operatorname{sen}(\pi/2M)} \qquad (17)$$

y por lo tanto, $B_N \to B$ conforme $M \to \infty$.

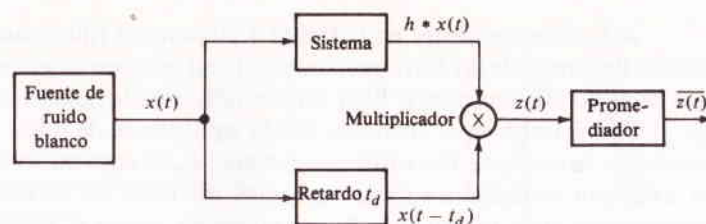



FIGURA 3.21
Medición de la respuesta al impulso, empleando ruido blanco.

Ruido blanco y mediciones de filtro ★

Dado que el ruido blanco contiene todas las frecuencias en igual proporción, resulta una señal conveniente para mediciones de filtros y para trabajo de diseño experimental. En consecuencia, las fuentes de ruido blanco con densidad de potencia calibrada han venido a ser instrumentos de laboratorio estándar. A continuación se analizarán algunas de las mediciones que se pueden hacer con estas fuentes.

Ancho de banda equivalente de ruido Supóngase que se conoce la ganancia de un amplificador y se desea conocer su ancho de banda equivalente de ruido. Para hacerlo, se aplica ruido blanco a la entrada y se mide la potencia de salida promedio con un medidor cuya respuesta de frecuencia sea constante de manera esencial dentro de la banda de paso del amplificador. El ancho de banda del ruido en cuestión será entonces, de la ecuación (14), $B_N = N/\mathcal{G}\eta$.

Respuesta de amplitud Para encontrar la respuesta de amplitud de un filtro dado, se aplica ruido blanco a la entrada de manera que el espectro de potencia de salida sea proporcional a $|H(f)^2|$. Entonces se explora el espectro de salida con un filtro pasabanda sintonizable cuyo ancho de banda sea constante y pequeño en comparación con las variaciones de $|H(f)|^2$. Así, si el filtro de exploración se centra en f_c, el voltaje de ruido rms en su salida será proporcional a $|H(f_c)|$. Variando f_c, se obtiene, punto por punto, la gráfica de $|H(f)|$.

Respuesta al impulso La figura 3.21 muestra un método para la medición de la respuesta al impulso $h(t)$ de un sistema dado. El instrumental que se requiere es una fuente de ruido blanco, un retardador de tiempo variable, un multiplicador y un dispositivo promediante.

Denotando al ruido de entrada como $x(t)$, la salida del sistema será $h * x(t)$, y la señal retardada $x(t - t_d)$. Así, la salida del multiplicador será

$$z(t) = x(t - t_d)[h * x(t)]$$

$$= x(t - t_d) \int_{-\infty}^{\infty} h(\lambda)x(t - \lambda)\, d\lambda$$

Ahora, supóngase que se promedia $z(t)$ sobre un tiempo suficientemente largo para obtener $\langle z(t)\rangle$. Si la fuente de ruido es ergódica y el sistema es lineal e invariante en el tiempo, la potencia promedio es el promedio de ensamble

$$\overline{z(t)} = \int_{-\infty}^{\infty} h(\lambda)\mathbf{E}[x(t - t_d)x(t - \lambda)]\, d\lambda$$

donde se han intercambiado los órdenes de integración y de promediado. Examinada la integral, se puede ver que

$$F(x) = \begin{cases} 0 & x < 0 \\ Kx^2 & 0 \le x \le 10 \\ 100K & x > 10 \end{cases}$$

Pero, con $x(t)$ que es el ruido blanco, la ecuación (6*b*) dice que $R_x(\lambda - t_d) = (\eta/2)\delta(\lambda - t_d)$. En consecuencia

$$R_v(\tau) \approx f_0 G_v(0) + 2f_0 \sum_{m=1}^{\infty} G_v(mf_0) \cos m\omega_0 \tau$$

Por lo tanto, $h(t)$ se mide por medio de la variación del retardo del tiempo t_d.

El lector puede cuestionar lo práctico de este método, en forma particular, puesto que $h(t)$ se puede obtener inmediatamente con la aplicación de un impulso (o un pulso breve e intenso) al sistema. En tanto esta sea una conclusión válida para muchas mediciones de *filtros*, hay muchos sistemas, tales como procesamiento industrial y sistemas de control, para los cuales no se puede dar una entrada impulsiva, o si se hace, podría dañar o destruir al sistema.

3.7 PROBLEMAS

3.1 (Secc. 3.1) Una moneda gruesa, pero legal, tiene la probabilidad de caer sobre su borde $P(E) = 0.2$. Evalúese lo siguiente si se lanza al aire dos veces: $P(EE)$, $P(\text{no } E)$, $P(\text{no } HH)$, $P(\text{pareja})$.
Respuesta: 0.04, 0.64, 0.84, 0.32.

3.2 (Secc. 3.1) Se sacan tres bolas en forma aleatoria de una urna que contiene 500 bolas blancas, 300 bolas rojas y 200 bolas negras. Suponiendo que cada bola se sustituya antes de que se saque la siguiente, encuéntrense las probabilidades de los resultados siguientes: cada bola es de diferente color; dos o más bolas son del mismo color; no bolas blancas; dos o más bolas negras; dado que una bola es negra.

3.3 (Secc. 3.1) Un sistema de datos binario emplea dos símbolos **0** y **1**, transmitidos con probabilidades P_0 y P_1. Obedeciendo a errores de transmisión, se puede cambiar un **0** por un **1** en el receptor con una probabilidad P_0, y de manera similar para P_1. Obténganse expresiones para lo siguiente: el número aproximado de errores en una secuencia de $n \gg 1$ dígitos; la probabilidad de que haya ocurrido un error dado que se ha recibido un **0**.

3.4 (Secc. 3.2) Se lanzan tres monedas al aire y se designa con X el número total de caras. Encuéntrese $P(x_i)$ y grafíquese la función de densidad de probabilidad cuando $P(H) = 0.4$.

3.5 (Secc. 3.2) Se lanza un par de dados normales. Totalizando el número de puntos que aparecen, los resultados posibles son $x_j = 2, 3, \ldots, 12$. Encuéntrese $P(x_j)$ y grafíquese la función de densidad de probabilidad.

3.6 (Secc. 3.2) Una cierta variable aleatoria tiene

$$\overline{z(t)} = \frac{\eta}{2}\int_{-\infty}^{\infty} h(\lambda)\,\delta(\lambda - t_d)\,d\lambda = \frac{\eta}{2}\,h(t_d)$$

Encuéntrese K, evalúese $P(x \leq 5)$ y $P(5 < X \leq 7)$, y constrúyase gráfica de la función de densidad de probabilidad.

3.7 (Secc. 3.2) Dada $P(x) = a^2xe^{-ax}u(x)$, donde a es una constante, encuéntrese $F(x)$ y evalúese $P(X \leq 1/a$ y $P(1/a < X \leq 2/a)$.

3.8 (Secc. 3.2) Encuéntrense $P(X \leq 0)$, $P(0 < X \leq 1)$ y $P(X > 1)$ en el ejemplo 3.4.

3.9 (Secc. 3.2) Encuéntrense $P(0 < Z \leq 1)$, $P(Z \leq 1)$, y $P(Z \neq 0)$ en el ejercicio 3.4.

3.10 (Secc. 3.2) Sea $Z = 3X^{-5}$ donde $p_x(x)$ está dada por la ecuación (9*a*). Encuéntrese $F(z)$.

3.11 (Secc. 3.2) Emplee la ecuación (16) para demostrar que $\int_{-\infty}^{\infty} p(x/y)dx = 1$. Explíquese por qué esto debe ser verdadero.

3.12 (Secc. 3.3) Calcúlese la medida y la variancia de X cuando $p(x) = e^{-x}u(x)$.
Respuesta: $\bar{x} = \sigma_x^2 = 1$.

3.13 (Secc. 3.3) Calcúlese el valor medio y la variancia de X dada $p(x)$ en el problema 3.7.

3.14 (Secc. 3.3) Desarróllese $E[(x - \bar{x})^3$ en términos de los momentos de x.

3.15 (Secc. 3.3) Demuéstrese que $E[(x - c)^2]$ es mínima cuando $c = \bar{x}$.

3.16 (Secc. 3.3) En relación con el ejemplo 3.5, encuéntrese la gráfica $P(|X| \leq k\ 2)$ contra k, y compárese con la desigualdad de Chebyshev.

3.17 (Secc. 3.3) Dado que X y Y son estadísticamente independientes, encuéntrese σ_z^2 en términos de los parámetros estadísticos de X y Y cuando: $z = X - Y + c$, donde c es una constante; $Z = XY$ y $\bar{x} = \bar{y} = 0$.

3.18 (Secc. 3.3) Principiando con $E[(x - y)^2]$ obténgase un límite superior en $\overline{xy}$ cuando X y Y no son estadísticamente independientes.

3.19 (Secc. 3.3) Se hacen y promedian N observaciones independientes de una variable aleatoria para dar

$$\mathrm{E}[x(t - t_d)x(t - \lambda)] = R_x(\lambda - t_d)$$

donde X_j es el valor j-ésimo observado. Demuéstrese que $\bar{y} = \bar{x}$ y $\sigma_y = \sigma_x/\sqrt{N}$.

3.20 (Secc. 3.4) Aplicando la distribución binomial, encuéntrese la probabilidad de que habrá menos de tres caras al lanzar 10 monedas legales al aire. ¿Cuál es el valor esperado del número de caras?

3.21 (Secc. 3.4) ¡Cuan $m \gg 1$ y $\epsilon \gg 1$, la distribución binomial se puede aproximar por medio de la *distribución de Poisson* $P(n) = e^{-\bar{n}}\,\bar{n}^n/n!$, donde $\bar{n} = m\epsilon$. Úsese esto para encontrar la probabilidad de que no haya errores en 100 dígitos cuando $\epsilon = 10^{-3}$.

3.22 (Secc. 3.4) Usted ha diseñado un sistema de transmisión digital con una probabilidad de error de 10^{-6} por dígito. El cliente probará el sistema enviando un mensaje conocido de 10^{-6} dígitos y comparándolo con el mensaje recibido. Si hay más de dos errores, le despedirán a usted. Calcule la probabilidad de perder su trabajo. (*Sugerencia*: Véase el problema 3.21.)

3.23 (Secc. 3.4) Sea $y = \cos \pi X$ donde X tiene una función de densidad de probabilidad uniforme sobre $[-\frac{1}{2}, \frac{1}{2}]$. Encuéntrense $\bar{y}$ y σ_y.

3.24 (Secc. 3.4) Una variada gaussiana tiene $\bar{x} = 10$ y $\sigma_x = 20$. Encuéntrense $P(X > 20)$, $P(10 < X \leq 20)$, $P(0 < X \leq 20)$.
Respuesta: 0,31, 0.19, 0.38, 0.69.

3.25 (Secc. 3.4) Se sabe que un voltaje aleatorio es gaussiano con $\bar{x} = 0$ y $\overline{x^2} = 9$. Encuéntrese el valor de c tal que $|X| \leq c$ para:
(*a*) 90% del tiempo
(*b*) 99% del tiempo

3.26 (Secc. 3.4) Deduzca la aproximación, ecuación (10).

3.27★ (Secc. 3.4) Sea $Z = X_1 + 2X_2$ donde X_1 y X_2 son variadas gaussianas con $\overline{x_1^2} = \overline{x_2^2} = 25$, $m_1 = 0$, $m_2 = 3$, y $x_1, x_2 = 6$. Encuéntrese $p_z(z)$.

3.28★ (Secc. 3.4) Sea $X_2 = X_1^2$ tal que, de manera obvia, X_1 y X_2 no son independientes. Ello, no obstante, demuéstrese que son no correlacionadas ($\rho_{12} = 0$) si $p(x_1)$ tiene simetría par.

3.29 (Secc. 3.5) Considérese el proceso $v(t) = A\cos(\omega_0 t + \Theta)$, donde A y Θ son variables aleatorias independientes sobre el ensamble y Θ tiene una función de densidad de probabilidad uniforme sobre $[-\pi, \pi)$. Demuéstrese que $E[v(t)] = 0$ y $E[v^2(t)] = \bar{A}^2/2$. (*Sugerencia*: recuérdese que el promedio de ensamble se debe tomar con respecto a ambas variables aleatorias.)

3.30 (Secc. 3.5) Demuéstrese que el proceso del problema **3.29** no es ergódico.

3.31 (Secc. 3.5) Si $v(t)$ es un proceso ergódico aleatorio, periódico con periodo T_0, demuéstrese que $R_v(\tau)$ también tiene periodo T_0.

3.32 (Secc. 3.5) Una cierta señal aleatoria tiene una componente de CD de 2 volts (V) y un valor rms de 4 V. Mediciones posteriores indican que $v(t)$ y $v(t - \tau)$ son independientes para $|\tau| > 5$ (microsegundos), mientras que $R_v(\tau)$ disminuye linealmente con $|\tau|$ para $|\tau| \leq 5\ \mu$s.
(*a*) Haga gráfica de $R_v(\tau)$ y dimensiónese en forma completa.
(*b*) Encuéntrese y construya gráfica de $G_v(f)$.

3.33 (Secc. 3.5) Sea $z(t) = v(t) + v(t - T)$, donde $v(t)$ es una señal aleatoria y T es una constante. Encuéntrense $R_z(\tau)$ y $G_z(f)$ en términos de $R_v(\tau)$ y $G_v(f)$.

3.34 (Secc. 3.5) Supóngase que $v(t)$ en la ecuación (11*a*) es una onda binaria aleatoria (ejemplo 3.7). Bosquéjense $R_z(\tau)$ y $G_z(f)$ tomando $f_0 \gg 1/T$. Demuéstrese también que $E[z^2(t)] = \langle z^2(t)\rangle = A^2/2$.

3.35 (Secc. 3.5) Sea $z(t) = v(t)w(t)$, donde $v(t)$ y $w(t)$ son señales aleatorias independientes.
(*a*) Demuéstrense que $R_z(\tau) = R_v(\tau)R_w(\tau)$ y en consecuencia $G_z(f) = G_v * G_w(f)$.
(*b*) Empleénse estos resultados y el ejemplo 3.6 para deducir las ecuaciones (11*b*) y (11*c*).

3.36 (Secc. 3.5) Por medio de la inserción de la ecuación (17*a*) en la ecuación (2), demuéstrese que

$$Y = \frac{1}{N}\sum_{j=1}^{N} X_j$$

3.37 (Secc. 3.6) Un resistor metálico de 1 kiloohm (kΩ) está a 29 K. Calcúlese el voltaje de ruido rms a circuito abierto; estímese el intervalo de frecuencia sobre el cual $G_v(f)$ es de manera esencialmente constante; evalúense $G_v(f)$, $G_i(f)$, y $G_a(f)$ en este intervalo, especifiquense sus unidades.

3.38 (Secc. 3.6) Dos resistores térmicos R_1 y R_2 a temperaturas $\mathscr{T}_1$ y $\mathscr{T}_2$ están conectados en serie. Encuéntrese la temperatura de ruido $\mathscr{T}_N$ de la combinación. (*Sugerencia*: Empléese

el modelo de Thevenin, recordando que las densidades espectrales se suman cuando las fuentes son independientes.)
Respuesta: $(R_1\mathcal{T}_1 + R_2\mathcal{T}_2)/(R_1 + R_2)$.

3.39 (Secc. 3.6) Resuélvase el problema 3.38 para una conexión en paralelo.

3.40 (Secc. 3.6) Un diodo al vacío de temperatura limitada generador de ruido tiene $G_v(f) = 2Rk\mathcal{T}_0 + eI_bR^2$, donde R es la resistencia de la fuente, e es la carga electrónica, e I_b la corriente de CD del diodo. Obténgase una expresión para la temperatura de ruido y evalúesele cuando $R = 50\ \Omega$ y $I_b = 20$ miliamperes (mA).

3.41 (Secc. 3.6) La entrada a una retención de orden cero, es ruido blanco, ejemplo 2.13, sección 2.5. Encuéntrese $G_y(f)$, $R_y(\tau)$ y $\overline{y^2}$.

3.42 (Secc. 3.6) El ruido que proviene de un resistor de 10 kΩ a la temperatura ambiente pasa a través de un filtro pasabajas ideal con $B = 2.5$ MHz y ganancia unitaria. Obténgase una expresión para la función de densidad de probabilidad del voltaje de salida $y(t)$.
Respuesta: Gaussiana con $\overline{y} = 0$, $\sigma_y = 2 \times 10^{-5}$.

3.43 (Secc. 3.6) El ruido filtrado del problema 3.42 pasa a través de un rectificador de onda completa, dando $z(t) = |y(t)|$. Encuéntrese $p_z(z)$ y evalúense $\overline{z}$ y σ_z.

3.44★ (Secc. 3.6) El ruido filtrado del problema 3.42 se muestrea cada 0.1 microsegundos (μs). Denotando dos valores muestra adyacentes como y_1 y y_2, obténgase una expresión para sus funciones de densidad de probabilidad conjunta $p(y_1, y_2)$. (*Sugerencia*: La covarianza σ_{12} se puede encontrar a partir de $R_y(\tau)$.)

3.45 (Secc. 3.6) Encuéntrese el ancho de banda equivalente de ruido B_N para el filtro pasabajas gaussiano del problema 2.44 y compáresele con el ancho de banda a 3 dB.

4

COMUNICACIÓN DE BANDA BASE

La comunicación de banda base se refiere a la transmisión de señal *sin modulación*; el nombre proviene del hecho de que la transmisión de banda base no incluye la traslación de frecuencia del espectro del mensaje que caracteriza a la modulación. Y aunque la mayoría de los sistemas de comunicación son sistemas de modulación, la transmisión de banda base merece nuestro estudio. Por una razón, los enlaces conectados con la banda base son parte de muchos sistemas de modulación; por otra, muchos de los conceptos y parámetros de la comunicación de banda base conducen directamente a la modulación. Pero quizá la más importante reside en las características de elaboración de la transmisión de banda base que sirve como normas útiles cuando se comparan los diferentes tipos de modulación.

Este capítulo, por lo tanto, está dedicado a la comunicación de banda base. Empieza con una investigación de las dos limitaciones fundamentales de la señalización eléctrica mencionadas en el capítulo 1, es decir, el ruido y el ancho de banda. Los resultados se aplican después a tres clases distintas de transmisión de banda base: analógica, de pulsos y digital. La cobertura abarca desde los cálculos de diseño elementales pero significativos, hasta el análisis opcional de técnicas complejas mejoradas.

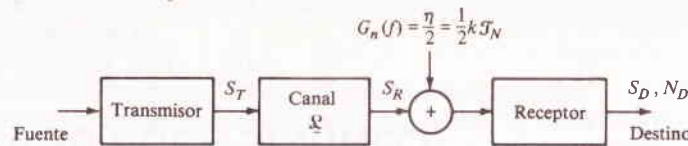



FIGURA 4.1
Parámetros de un sistema de comunicación.

4.1 SEÑALES Y RUIDO

Se observó en el capítulo 3 que el ruido resulta ser un factor significativo en la comunicación eléctrica, cuando la señal recibida es muy débil y por ello del mismo orden de magnitud que el siempre presente ruido térmico. A su vez, el nivel tan pequeño de la señal recibida se debe en primer lugar a la gran cantidad de pérdida de potencia que caracteriza a la transmisión a larga distancia. En consecuencia, este análisis de señales y ruido empieza considerando las pérdidas de transmisión. La figura 4.1 pone este tema en el contexto y señala los diferentes parámetros del sistema con los cuales se tratará. Estos parámetros se definen como sigue:

S_T = potencia de la señal a la salida del transmisor
$\mathcal{L}$ = pérdida de potencia de transmisión del canal
S_R = potencia de la señal a la entrada del receptor
$\mathcal{T}_N$ = temperatura de ruido en relación con la entrada del receptor
η = densidad de ruido (supuesta constante) a la entrada del receptor
S_D = potencia de la señal en el destino
N_D = potencia de ruido en el destino

Este conjunto notacional se emplea en lo que resta del texto y el lector debe estudiar en forma cuidadosa las definiciones.

Pérdidas de transmisión

Cuando se dice que un amplificador tiene *ganancia de potencia* $\mathcal{G}$, significa que las potencias de entrada y salida están relacionadas por $P_{\text{salida}} = \mathcal{G}\, P_{\text{entrada}}$, así,

$$\mathcal{G} \triangleq \frac{P_{\text{salida}}}{P_{\text{entrada}}} \tag{1a}$$

Con frecuencia la ganancia se expresa en decibeles† como

$$\mathcal{G}_{\text{dB}} \triangleq 10 \log_{10} \mathcal{G} = 10 \log_{10} \frac{P_{\text{salida}}}{P_{\text{entrada}}} \tag{1b}$$

† Véase la Tabla E para conversiones y manejo de decibeles.

En consecuencia, si $P_{\text{salida}} = 100\, P_{\text{entrada}}$, $\mathcal{G} = 100$ y $\mathcal{G}_{dB} = 20$ dB; si la ganancia de potencia es la unidad, entonces $\mathcal{G}_{\text{dB}} = 0$ dB. Pero casi todos los canales de transmisión tienen $P_{\text{salida}} < P_{\text{entrada}}$ y $\mathcal{G} < 1$. Así , se define la ***pérdida de potencia,***

$$\mathcal{L} \triangleq \frac{1}{\mathcal{G}} = \frac{P_{\text{entrada}}}{P_{\text{salida}}} \tag{2a}$$

$$\mathcal{L}_{\text{dB}} = -\mathcal{G}_{\text{dB}} = 10 \log_{10} \frac{P_{\text{entrada}}}{P_{\text{salida}}} \tag{2b}$$

también conocida como la *atenuación*.

Aunque se han definido $\mathcal{G}$ y $\mathcal{L}$ como relaciones de potencia, se pueden expresar como relaciones de voltajes o corrientes cuadráticos medios, por ejemplo,

$$\mathcal{G} = \frac{\overline{v_{\text{salida}}^2}}{\overline{v_{\text{entrada}}^2}} \tag{3}$$

la barra se pone para uno u otro caso: el promedio de tiempo o el promedio de ensamble, según convenga. La ecuación (3) es consistente con el uso de la potencia normalizada, y no será necesario hacer ditinciones entre ganancia de potencia disponible, ganancia de potencia de transductor, etc.

Para aclarar la ecuación (2), las relaciones de entrada y salida para líneas de transmisión, cables coaxiales y guías de onda, son todas de la forma

$$P_{\text{salida}} = e^{-2\alpha l} P_{\text{entrada}} = 10^{-8.68\alpha l} P_{\text{entrada}} \tag{4}$$

donde α es el coeficiente de atenuación y l es la longitud de la trayectoria, por lo tanto, la pérdida en decibeles es

$$\mathcal{L}_{\text{dB}} = 8.68\alpha l \tag{5}$$

En la Tabla 4.1 se listan algunos valores representativos. En realidad, las ecuaciones (4) y (5) son válidas de manera estricta sólo para señales senoidales y α depende de la frecuencia, como se desprende de la tabla. Por razones de tiempo no se trata la dependencia de la frecuencia, puesto que el punto a tratar aquí es lo relacionado con los potencialmente

Tabla 4.1 VALORES TÍPICOS DE PÉRDIDAS DE TRANSMISIÓN

Medio de transmisión	Frecuencia	Pérdidas, dB/km
Línea aérea de dos conductores (0.3 cm de diámetro)	1 kHz	0.05
Par trenzado (calibre 16)	10 kHz	2
	100 kHz	3
	300 kHz	6
Cable coaxil (1 cm de diámetro)	100 kHz	1
	1 MHz	2
	3 MHz	4
Cable coaxil (15 cm de diámetro)	100 MHz	1.5
Guía de onda rectangular (5 × 2.5 cm)	10 GHz	5
Guía de onda helicoidal (5 cm de diámetro)	100 GHz	1.5

grandes valores de pérdidas. Para subrayar este punto: una línea formada por un par de alambres torcidos de calibre 16 y de 20 km de longitud, a 100 kHz tiene $\mathscr{L}_{db} = 3 \times 20 = 60$ dB o $\mathscr{L} = 10^6$, lo cual significa que ¡$P_{salida} = P_{entrada}/\mathscr{L} = 10^6 P_{entrada}$! Obsérvese también que la pérdida en decibeles es directamente proporcional a l; así, continuando con el ejemplo, una línea de 40 km tiene ¡ $\mathscr{L}_{dB} = 2 \times 60 = 120$ dB y $P_{salida} = 10^{12} P_{entrada}$!

En vista de los resultados anteriores, no es sorprendente que la transmisión por radio propagación sea a menudo preferida para distancias grandes. Aunque los sistemas de radio incluyen la modulación, que será tratada en subsecuentes capítulos, parece apropiado mencionar ahora la diferencia entre la transmisión por cable y por radio, al menos en lo que a pérdida de trayectoria se refiere.

De manera específica, la relación de potencias en una trayectoria de línea de vista es

$$\frac{P_{salida}}{P_{entrada}} = \mathscr{G}_{TA}\mathscr{G}_{RA}\left(\frac{\lambda}{4\pi l}\right)^2 \qquad (6)$$

donde $\mathscr{G}_{TA}$ y $\mathscr{G}_{RA}$ son las ganancias de potencia de las antenas transmisora y receptora y λ es la longitud de onda de la portadora. Así,

$$\mathscr{L}_{dB} = 22 + 10\log_{10}\left(\frac{l}{\lambda}\right)^2 - (\mathscr{G}_{TA_{dB}} + \mathscr{G}_{RA_{dB}}) \qquad (7)$$

de tal manera que una trayectoria de 20 km con $\lambda = 1$ metro y $\mathscr{G}_{TA} = \mathscr{G}_{RA} = 16$ dB tiene $\mathscr{L}_{dB} = 22 + 86 - 32 = 76$ dB. Pero nótese que la ecuación (7) *no* es proporcional a l; en efecto, doblando a l se incrementa la pérdida en sólo 6 dB (afirmación que el lector debe confirmar por sí mismo). Por lo tanto, una trayectoria de 40 km con los parámetros anotados tiene $\mathscr{L}_{dB} = 76 + 6 = 82$ dB. De manera incidental, el límite superior para la línea de vista está alrededor de los 40 km sobre terreno plano, a menos que se empleen torres de antena más altas.

Volviendo a la figura 4.1, se puede expresar ahora S_R en términos de S_T y $\mathscr{L}$ de manera simple como

$$S_R = \frac{S_T}{\mathscr{L}} \qquad (8)$$

Entonces, si la ganancia de potencia del receptor es $\mathscr{G}_R$,

$$S_D = \mathscr{G}_R S_R = \frac{\mathscr{G}_R}{\mathscr{L}} S_T \qquad (9a)$$

Ecuaciones como la (9*a*) pueden escribirse en la forma

$$S_{D_{dBW}} = \mathscr{G}_{R_{dB}} - \mathscr{L}_{dB} + S_{T_{dBW}} \qquad (9b)$$

donde dBW queda establecido para decibeles arriba de 1 watt, por ejemplo, 20 dBW = 100 W. En consecuencia, la multiplicación y la división han sido sustituidas por la suma y la resta de valores en decibeles, lo cual obedece a que son muy empleados por los ingenieros en comunicaciones. Los índices adicionales empleados en la ecuación (9*b*), etc., son de uso molesto y se omitirán en lo sucesivo; esto debe sobreentenderse en el contexto, cuando de valores en decibeles se trate.

Ruido aditivo

Habiendo entregado la señal al receptor, véanse los efectos del ruido contaminante. La figura 4.2 considera sólo la porción conveniente de la figura 4.1. La forma de onda en la entrada del receptor es $K_R x(t)$, donde $x(t)$ es la señal que conduce la información, a la cual se le agrega el ruido blanco $G_n(f) = \eta/2 = k\mathscr{T}_N/2$. La temperatura de ruido $\mathscr{T}_N$ representa a *todo* el ruido del sistema, en relación con la entrada del receptor †. Los valores numéricos pueden abarcar desde alrededor de 60°K en un sistema de bajo ruido diseñado en forma cuidadosa hasta varios miles de grados.

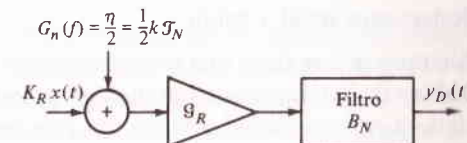



FIGURA 4.2
Modelo de receptor con ruido blanco aditivo.

Para propósitos de análisis, se ha dividido al receptor en dos partes: un amplificador con ganancia de potencia $\mathscr{G}_R$ (y ganancia de voltaje, $\sqrt{\mathscr{G}_R}$) seguido de un filtro con ancho de banda equivalente de ruido B_N y $\mathscr{G} = 1$. La función del filtro es dejar pasar a $x(t)$ y rechazar tanto ruido como sea posible, es decir, aquellos componentes de ruido fuera del intervalo de frecuencia de la señal, en otras palabras, el ruido *fuera de banda.*

Suponiendo que no existen no linealidades en el receptor, la salida total es

$$y_D(t) = \sqrt{\mathscr{G}_R}\, K_R x(t) + n_D(t) \qquad (10)$$

donde $n_D(t)$ es el ruido de salida. No es irrazonable suponer más allá de que $x(t)$ y $n_D(t)$ son *estadísticamente independientes* y que $\overline{n}_D = 0$; bajo estas condiciones, $x(t)$ y $n_D(t)$ son *incoherentes* y sus correlaciones cruzadas son cero, es decir, $R_{xn_D}(\tau) = R_{n_Dx}(\tau) = 0$. Por lo tanto, empleando la ecuación (10), sección 3.5,

$$R_{y_D}(\tau) = \mathscr{G}_R K_R^2 R_x(\tau) + R_{n_D}(\tau) \qquad (11)$$

y

$$\overline{y_D^2} = \underbrace{\mathscr{G}_R K_R^2 \overline{x^2}}_{S_D} + \underbrace{\overline{n_D^2}}_{N_D} \qquad (12)$$

† Véase el Apéndice B para el método.

EJERCICIO 4.1 Se va a transmitir una señal a 40 km, empleando una línea de transmisión cuya pérdida es 3 dB/km; el receptor tiene $\mathcal{T}_N = 10\mathcal{T}_0$ y $B_N = 5$ kHz. Calcúlese ηB_N en decibeles sobre un watt (dBW) y encuéntrese el valor de S_T (en watts) que se requieren para obtener $(S/N)_D = 50$ dB. Repítase el segundo cálculo cuando existe un repetidor a la mitad de la trayectoria. *Respuesta*: $-$ 157 dBW, 20W, 40 μW.

4.2 DISTORSIÓN DE LA SEÑAL EN LA TRANSMISIÓN

Además del ruido, la otra limitación fundamental en la comunicación eléctrica es el ancho de banda, el ancho de banda finito de cualesquiera sistemas reales produce distorsión de la señal. Pero la *transmisión sin distorsión* no implica de manera necesaria que la salida sea idéntica a la entrada. Se pueden tolerar ciertas diferencias sin que se clasifiquen como distorsión. Nuestro propósito aquí es establecer el significado de la transmisión sin distorsión y los requisitos para ello. Con esta base, se pueden definir los diferentes tipos de distorsión e investigar sus efectos. Se recalcarán estos aspectos pertinentes a los sistemas de comunicación.

Establecido de manera burda, en la transmisión sin distorsión la salida debe "lucir" como la entrada. De manera más precisa, dada una señal de entrada $x(t)$, se dice que la salida es sin distorsión si difiere de la entrada sólo en una constante multiplicadora y en un retardo finito de tiempo. En forma analítica, se tiene transmisión sin distorsión si

$$y(t) = Kx(t - t_d) \tag{1}$$

donde K y t_d son constantes.

Las propiedades de una red sin distorsión se encuentran con facilidad examinando el espectro de salida

$$Y(f) = \mathcal{F}[y(t)] = Ke^{-j\omega t_d}X(f)$$

Ahora, por la definición de función de transferencia, $y(f) = H(f)X(f)$, tal que

$$H(f) = Ke^{-j\omega t_d} \tag{2a}$$

En palabras, una red en la que se tiene transmisión sin distorsión debe tener *respuesta de amplitud constante* y *corrimiento de fase lineal* negativo, o sea

$$|H(f)| = K \qquad \arg[H(f)] = -2\pi t_d f \pm m180° \tag{2b}$$

Se ha agregado el término $\pm\ m$ 180° para calcular los valores positivos o negativos de la constante. La fase cero es permisible, puesto que ello implica cero retardo en el tiempo. Se puede agregar una cualidad más a la ecuación (2): se requieren estas condiciones sólo en aquellas frecuencias para las cuales la señal de entrada tiene espectro diferente de cero. Así, si $x(t)$ es de banda limitada en W, la ecuación (2) debe satisfacerse sólo para $|f| < W$.

En la práctica la transmisión sin distorsión es una condición estricta, la cual, en el mejor de los casos, sólo puede ser satisfecha en forma aproximada. Es un hecho ine-

vitable en la transmisión de señales que la distorsión ocurrirá, aunque se puede reducir con un diseño apropiado. Por lo tanto, se atenderá al *grado* de distorsión, medido en alguna modalidad cuantitativa. De manera desafortunada, se ha comprobado que las medidas cuantitativas son más bien pesadas e imprácticas para propósitos de ingeniería. Como una aproximación alterna, la distorsión se ha clasificado conforme a tipos, y cada tipo considerado en forma separada. Pero antes de analizar los diferentes tipos, se debe recalcar que la *distorsión es distorsión*; una salida fuertemente distorsionada diferirá en forma significativa de la entrada, sin tomar en consideración la causa específica.

Las tres clasificaciones principales de distorsión son:

1 Distorsión de amplitud: $|H(f)| \neq K$
2 Distorsión de fase (retardo): arg. $[H(f)] \neq -2\pi t_d f \pm m$ 180°
3 Distorsión no lineal

Los primeros dos casos están en la categoría de la distorsión *lineal*. En el tercer caso el sistema incluye elementos no lineales, y su función de transferencia no está definida. A continuación se examinan éstos en forma individual.

Distorsión de amplitud

La distorsión de amplitud se describe con facilidad en el dominio de la frecuencia; significa en forma simple que las componentes de frecuencia de salida no están en la proporción correcta. Como la causa de esto es que $|H(f)|$ no es constante con la frecuencia, a la distorsión de amplitud se le designa algunas veces como *distorsión de frecuencia*.

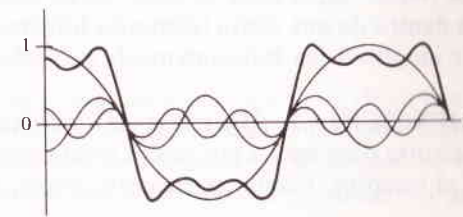



FIGURA 4.4
Señal de prueba $x(t) = \cos\omega_0 t - ⅓\cos 3\omega_0 t + 1/5\cos 5\omega_0 t$.

Las formas más comunes de distorsión de amplitud son atenuación excesiva o levantamiento de los extremos de las altas o bajas frecuencias en el espectro de la señal. Menos común, pero igualmente molesta, es una respuesta desproporcionada a una banda de frecuencia dentro del espectro. Mientras que la descripción del dominio de la frecuencia es fácil, los efectos en el dominio del tiempo son mucho menos obvios, salvo para señales muy simples. Como ilustración, una señal de prueba de manera conveniente simple es $x(t) = \cos\omega_0 t - ⅓\cos 3\omega_0 t + 1/5\cos 5\omega_0 t$, figura 4.4, una aproximación burda a una onda cuadrada. Si la componente de baja o alta frecuencia se atenúa a la mitad, las salidas resultantes son como se muestran en la figura 4.5. Como es de esperarse, la pérdida de los términos de alta frecuencia reduce la "agudeza" de la forma de onda.

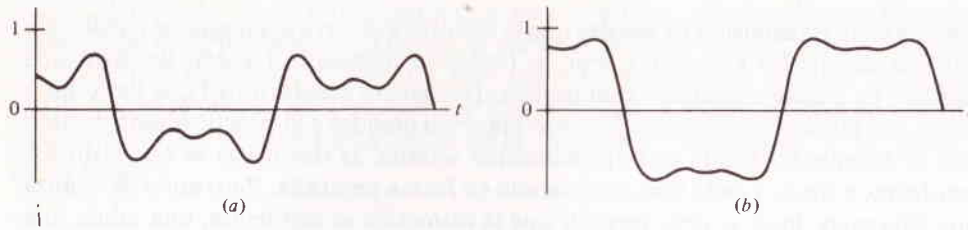



FIGURA 4.5
Señal de prueba con distorsión de amplitud. (*a*) Atenuada la baja frecuencia; (*b*) atenuada la alta frecuencia.

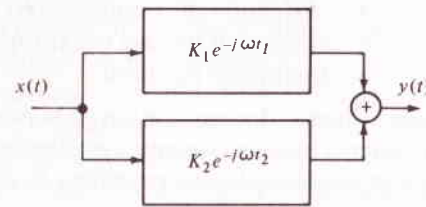



FIGURA 4.6
Diagrama a bloques de un sistema de radio con dos trayectorias de propagación.

Más allá de observaciones cualitativas, se puede decir un poco más acerca de la distorsión de amplitud sin un estudio experimental de tipos específicos de señales. Los resultados de tales estudios descansan por lo general en la *respuesta de frecuencia* requerida, es decir, el intervalo de frecuencias en el cual $|H(f)|$ debe ser constante dentro de una cierta tolerancia (digamos $\pm$ 1 dB) de tal manera que la distorsión de amplitud sea suficientemente pequeña.

EJERCICIO 4.2 Algunas veces los sistemas de radio presentan *distorsión multitrayectoria* ocasionada por dos (o más) trayectorias de propagación entre el transmisor y el receptor. Como un ejemplo simple, supóngase que la señal recibida es

$$y(t) = K_1 x(t - t_1) + K_2 x(t - t_2) \tag{3a}$$

Demuéstrese que la figura 4.6 es el diagrama a bloques equivalente y que, si $(K_2/K_1)^2 \ll 1$.

$$|H(f)| \approx K_1\left[1 + \frac{K_2}{K_1}\cos 2\pi f(t_2 - t_1)\right] \tag{3b}$$

Por lo que, una reflexión ''débil'' ocasiona *rizos* en la relación de amplitudes.

Corrimiento de fase y distorsión por retardo

Un corrimiento de fase lineal ocasiona un retardo de tiempo constante para todas las componentes de frecuencia de la señal. Así, con una respuesta de amplitud constante, se tiene una salida sin distorsión. Si el corrimiento de fase no es lineal, las diferentes componentes de frecuencia sufren diferentes retardos de tiempo, y a la distorsión resultante se le designa *distorsión de fase o distorsión por retardo*.

Para un corrimiento de fase arbitrario, el retardo en tiempo es una función de la frecuencia, llamado $t_d(f)$, y se puede encontrar escribiendo $[H(f)] = -2\pi f t_d(f)$, tal que

$$t_d(f) = -\frac{\arg\,[H(f)]}{2\pi f} \tag{4}$$

lo cual es independiente de la frecuencia sólo si $[H(f)]$ es lineal con ella.

Un área común de confusión es el *retardo en tiempo constante* contra *corrimiento de fase constante*. Lo primero es deseable y se requiere para transmisión sin distorsión. El segundo, en general, causa distorsión. Supóngase un sistema con corrimiento de fase constante θ. Entonces

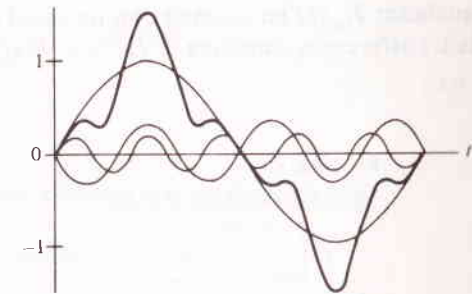



FIGURA 4.7
Señal de prueba con corrimiento de fase constante de $-90°$.

cada componente de frecuencia de la señal será retardado en $\theta/2\pi$ *ciclos* de su propia frecuencia; éste es el significado de corrimiento de fase constante. Pero los retardos en el tiempo serán diferentes, las componentes de frecuencia se mezclarán en el tiempo, y resultará la distorsión. Sin embargo, son aceptables los corrimientos de fase constantes $\theta = 0$ y $\pm\, m\ 180°$.

Aquel corrimiento de fase constante que produce distorsión se aclara de manera simple, volviendo a la señal de prueba de la figura 4.4 y desplazando cada componente en un cuarto de ciclo, $\theta = -90°$. Mientras que la entrada fue de manera burda una onda cuadrada, la salida se verá como una onda triangular, figura 4.7. Con un corrimiento de fase no lineal arbitrario, la deformación de la forma de onda puede ser aún más marcada.

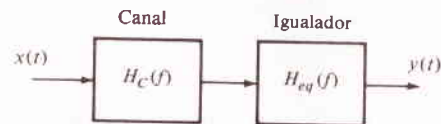
Se debe notar también en la figura 4.7 que los valores de *pico* de la señal desplazada en fase son en forma sustancial mayores (alrededor del 50%) que los de la señal de prueba de entrada. Esto no es debido a la respuesta de amplitud, puesto que las amplitudes de salida de los tres componentes de frecuencia de hecho no han cambiado; más aún, esto es así porque las componentes de la señal distorsionada alcanzan valores máximos o mínimos al mismo tiempo, lo cual no fue así a la entrada.

A la inversa, si se hubiera empezado con la figura 4.7 como la señal de prueba, un corrimiento de fase constante de + 90° conduciría a la figura 4.4 para la forma de onda de salida. Así, se ve que la *distorsión por retardo solamente* puede provocar un aumento o una disminución de los valores pico, y otras alteraciones en la forma de onda.

Es obvio que la distorsión por retardo puede ser crítica en la transmisión de pulsos, y se desperdicia mucho trabajo *igualando* el retardo en la transmisión para sistemas digitales de datos y casos semejantes. Por otra parte, el oído humano es en forma curiosa insensible a la distorsión por retardo; las formas de onda de las figuras 4.4 y 4.7 sonarían casi iguales al impulsar un altoparlante. Así, la distorsión por retardo es raro que se presente en la transmisión de voz o música.

Igualación

La distorsión lineal —es decir, la distorsión de amplitud y por retardo— es en forma teórica remediable con el uso de redes de *igualación*. La figura 4.8 muestra un igualador $H_{eq}(f)$ en cascada con un canal de distorsión $H_c(f)$. Puesto que la función de transferencia completa es $H(f) = H_c(f)H_{eq}(f)$

FIGURA 4.8
Canal con igualador para distorsión lineal.

La salida final será sin distorsión si $H_c(f)H_{eq}(f) = Ke^{-j\omega t_d}$, donde K y t_d son constantes más o menos arbitrarias. Por lo tanto, se requiere que

$$H_{eq}(f) = \frac{Ke^{-j\omega t_d}}{H_C(f)} \tag{5}$$

dondequiera que $X(f) \neq 0$.

Es raro el caso en que se pueda diseñar un igualador para satisfacer la ecuación (5) de manera exacta, lo cual explica por qué se dice que la igualación es un remedio teórico. Pero a menudo son posibles, aproximaciones excelentes, de tal manera que se puede reducir la distorsión lineal a un nivel tolerable. Es probable que la técnica de igualación más antigua sea el empleo de *bobinas de carga* en líneas telefónicas de pares torcidos†. Estas bobinas son inductores concentrados que se conectan en derivación a través de la línea cada kilómetro o más, mejorando la relación de amplitudes típica, ilustrada en la figura 4.9. Se ha diseñado otros circuitos con elementos concentrados para tareas de igualación específicas.

En fecha más reciente el *igualador de línea de retardo variable* o *filtro transversal* ha surgido como un dispositivo conveniente y flexible para ilustrar el principio, la figura 4.10 muestra una línea de retardo con un retardo de tiempo total 2Δ que tiene derivaciones en cada extremo y en la parte central. Las salidas de las derivaciones pasan a través de ganancias ajustables, c_{-1}, c_0, y c_1, y sumadas conforman la salida final. Así

† En Everitt y Anner (1956, Cap. 8) se trata la teoría de cargado.

$$y(t) = c_{-1}x(t) + c_0 x(t-\Delta) + c_1 x(t-2\Delta) \tag{6a}$$

y

$$H_{eq}(f) = c_{-1} + c_0 e^{-j\omega\Delta} + c_1 e^{-j\omega 2\Delta}$$
$$= (c_{-1}e^{+j\omega\Delta} + c_0 + c_1 e^{-j\omega\Delta})e^{-j\omega\Delta} \tag{6b}$$

FIGURA 4.9
Relación de amplitudes de una línea telefónica típica con y sin bobinas de carga para igualación.

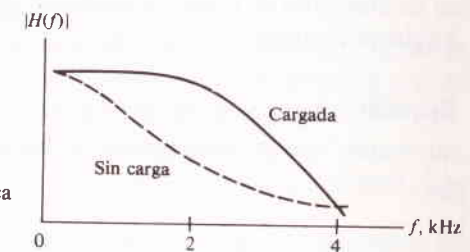
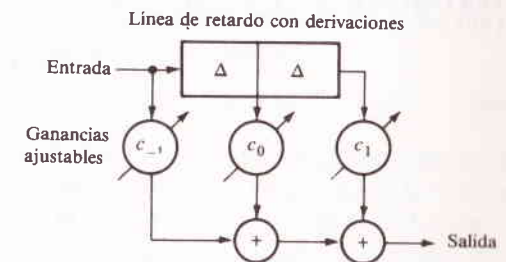



FIGURA 4.10
Igualador de línea de retardo con derivaciones (filtro transversal) con tres de ellas.

Como es obvio, éste es un arreglo conveniente, puesto que las ganancias de derivación varían con más facilidad que los elementos concentrados. Para demostrar la flexibilidad, supóngase $c_{-1} = c_1 < c_0/2$; entonces

$$|H_{eq}(f)| = c_0 + 2c_1 \cos \omega\Delta \qquad \arg [H_{eq}(f)] = -\omega\Delta$$

Por otra parte, si $c_{-1} = -c_1$ y $|c_1| \ll c_0$, entonces

$$|H_{eq}(f)| \approx c_0 \qquad \arg [H_{eq}(f)] \approx -\omega\Delta - \frac{2c_1}{c_0} \operatorname{sen} \omega\Delta$$

Por lo tanto, dependiendo de las ganancias de derivación, se pueden igualar los rizos de amplitud o rizos de fase, o ambos.

Generalizando la ecuación (6*b*) al caso de una línea de retardo de $2M\Delta$ con $2M + 1$ derivaciones,

$$H_{eq}(f) = \left(\sum_{m=-M}^{M} c_m e^{-j\omega m\Delta} \right) e^{-j\omega M\Delta} \tag{7}$$

la cual tiene la forma de una *serie exponencial de Fourier* con periodicidad de frecuencia $1/\Delta$. Por lo tanto, dado un canal $H_c(f)$ que va a ser igualado sobre $|f| < W$, se puede aproximar el segundo miembro de la ecuación (5) por medio de una serie de Fourier con periodicidad de frecuencia $1/\Delta \geq W$ (con lo cual se determina Δ), y estimar el número de términos significativos (con lo cual se determina M), y acoplar las ganancias de derivación a los coeficientes de la serie. Por lo tanto, este método teórico de elevación de potencia puede no necesitarse en casos simples tales como el del siguiente ejemplo.

Ejemplo 4.1

Supóngase que se desea igualar la distorsión multitrayectoria descrita en el ejercicio 4.2, donde

$$H_C(f) = K_1 \left[1 + \frac{K_2}{K_1} e^{-j\omega(t_2 - t_1)} \right] e^{-j\omega t_1}$$

con $(K_2/K_1)^2 \ll 1$ y $t_2 > t_1$. Aplicando la ecuación (5), el igualador debe tener

$$H_{eq}(f) = \frac{K}{K_1} \frac{e^{-j\omega(t_d - t_1)}}{1 + (K_2/K_1)e^{-j\omega(t_2 - t_1)}}$$

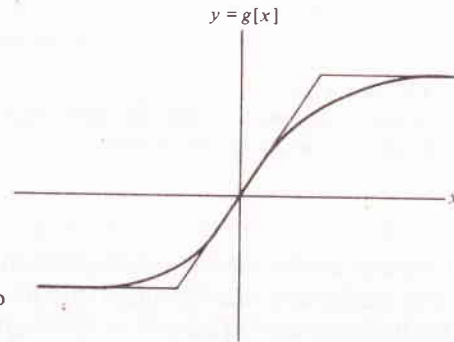



FIGURA 4.11
Característica de transferencia de un dispositivo no lineal.

Tomando $K = K_1$ y $t_d = t_1$ y desarrollando el denominador como una serie binomial de tres términos

$$H_{eq}(f) \approx 1 - \frac{K_2}{K_1} e^{-j\omega(t_2 - t_1)} + \left(\frac{K_2}{K_1}\right)^2 e^{-j\omega(t_2 - t_1)} e^{-j\omega(t_2 - t_1)}$$

$$\approx \left[e^{+j\omega(t_2 - t_1)} - \frac{K_2}{K_1} + \left(\frac{K_2}{K_1}\right)^2 e^{-j\omega(t_2 - t_1)} \right] e^{-j\omega(t_2 - t_1)}$$

Comparando esto con la ecuación (6*b*) se revela que un filtro transversal de tres derivaciones funcionarán si $c_{-1} = 1$, $c_0 = -(K_2/K_1)$, $c_1 = (K_2/K_1)^2$, y $\Delta = t_2 - t_1$.

Distorsión no lineal

Un sistema que tiene elementos no lineales no se puede describir por medio de una función de transferencia. En vez de eso, los valores instantáneos de entrada y salida se relacionan por medio de una curva o función $y(t) = g[x(t)]$, designada en forma común como la *característica de transferencia*. La figura 4.11 es una característica de transferencia representativa; el achatamiento de la salida para valores de entrada grandes es el efecto familiar de saturación y corte de los amplificadores a transistores. Se considerarán sólo dispositivos *sin memoria* para los cuales la característica de transferencia resulta una descripción completa.

Bajo condiciones de entrada de señal pequeña, es posible hacer lineal la característica de transferencia en una modalidad a tramos, como se muestra con las líneas delgadas en la figura. La aproximación más general es una aproximación polinómica a la curva, de la forma

$$y(t) = a_1 x(t) + a_2 x^2(t) + a_3 x^3(t) + \cdots \tag{8a}$$

Son las más altas potencias de $x(t)$ en esta ecuación que dan lugar a la distorsión no lineal.

Aun cuando no se tiene la función de transferencia, se puede encontrar el espectro de salida, al menos de una manera formal, transformando la ecuación (8*a*). De manera específica recurriendo al teorema de la convolución,

$$Y(f) = a_1 X(f) + a_2 X * X(f) + a_3 X * X * X(f) + \cdots \tag{8b}$$

Ahora, si $x(t)$ es de banda limitada en W, la salida de una red lineal no contendrá frecuencias más allá de $|f| < W$. Pero en el caso no lineal, se ve que la salida incluye $X * X(f)$, la cual es de banda limitada en $2W$, $X * X * X(f)$, la cual es de banda limitada en $3W$, etc. Las no linealidades han dado lugar, por lo tanto, a componentes de frecuencia que no estaban en la entrada. Además, puesto que $X * X(f)$ puede contener componentes para $|f| < W$, esta porción del espectro se traslapa al de $X(f)$. Empleando técnicas de filtrado, los componentes agregados en $|f| > W$ se pueden remover, pero no hay una forma apropiada de rechazar las componentes agregadas en $|f| < W$. Éstas, de hecho, constituyen la distorsión no lineal.

Se obtiene una medida cuantitativa de la distorsión no lineal, empleando una onda cosenoidal simple, $x(t) = \cos \omega_0 t$, como la entrada. Insertando en la ecuación (8*a*) y desarrollando se llega a

$$y(t) = \left(\frac{a_2}{2} + \frac{3a_4}{8} + \cdots\right) + \left(a_1 + \frac{3a_3}{4} + \cdots\right) \cos \omega_0 t + \left(\frac{a_2}{2} + \frac{a_4}{4} + \cdots\right) \cos 2\omega_0 t + \cdots$$

Por lo que, la distorsión no lineal aparece como *armónicas* de la onda de entrada. La cantidad de distorsión por segunda armónica es la relación de la amplitud de este término a la de la fundamental, o en porcentaje:

$$\text{Distorsión por segunda armónica} = \frac{a_2/2 + a_4/4 + \cdots}{a_1 + 3a_3/4 + \cdots} \times 100\%$$

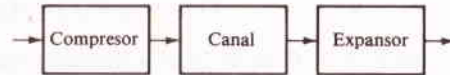
Las armónicas de orden superior se tratan de manera similar. Sin embargo, sus efectos son en forma general mucho menores, y muchas se pueden remover por completo por medio de filtrado.

Si la entrada es una suma de dos ondas cosenoidales, digamos cos $\omega_1 t$ + cos $\omega_2 t$, la salida incluirá todas las armónicas de f_1 y f_2, más los términos de producto de cruz con los que se llega a $f_2 - f_1, f_2 + f_1, f_2 - 2f_1$, etc. Estas frecuencias suma y diferencia se designan como *distorsión de intermodulación*. Generalizando el efecto de intermodulación, si $x(t) = x_1(t) + x_2(t)$, entonces $y(t)$ contiene el *producto cruz* $x_1(t)x_2(t)$ (y productos de orden superior, que se omiten aquí). En el dominio de la frecuencia $x_1(t)x_2(t)$ resulta ser $X_1 * X_2(f)$; y aun cuando $X_1(f)$ y $X_2(f)$ pueden ser separadas en frecuencia, $X_1 * X_2(f)$ pueden traslaparse sobre ambas, dando lugar a una forma de diafonía. Este aspecto de distorsión no lineal es de particular interés en sistemas de transmisión telefónica. Por otra parte, el término producto cruz es el resultado deseado cuando se emplean dispositivos no lineales para propósitos de modulación.

Compresión—Expansión (Compansión)

Aunque la distorsión no lineal no se puede remediar en una forma perfecta, se le puede reducir por medio de un diseño cuidadoso. La idea básica es estar seguro de que la señal no excede el intervalo de operación lineal de la característica de transferencia del canal. En forma irónica, una estrategia en relación con esto incluye

FIGURA 4.12
Sistema de compresión-expansión.

Compresor | Canal | Expansor

dos dispositivos *no lineales*, que son un compresor y un expansor, como se muestra en la figura 4.12.

Un compresor es un dispositivo que tiene mayor amplificación a bajos niveles de señal que a los altos. Puesto que la amplificación es la derivada de la característica de transferencia con respecto a la entrada, una característica $g_{comp}[x(t)]$ de compresor típica sería como la mostrada en la figura 4.13. Nótese que un compresor comprime el intervalo de la señal de salida. Por lo que, si el intervalo comprimido cae dentro del intervalo lineal del canal, la señal en la salida del canal será proporcional a $g_{comp}[x(t)]$ la cual se distorsiona por el compresor pero no el canal. En forma ideal, entonces, el expansor debe tener una característica que complemente en forma perfecta al compresor; es decir, menos amplificación a niveles bajos de señal, etc. Así, la salida final será proporcional a $g_{exp}\{g_{comp}[x(t)]\} = x(t)$ como se desea.

Es uso conjunto de la compresión y la expansión se conoce como compansión (¿sorpresa?) y resulta de un valor singular en sistemas telefónicos. Además de reme-

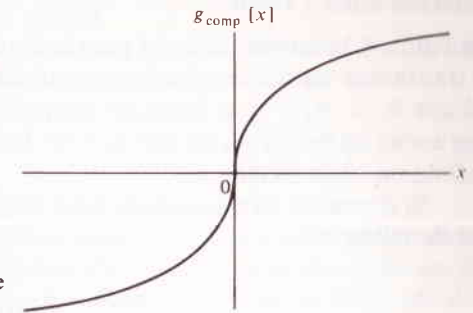



FIGURA 4.13
Característica de transferencia típica de un compresor.

diar la distorsión no lineal, la compansión tiende a compensar la diferencia de nivel de señales entre sonidos altos y suaves. En realidad, esto último es la ventaja clave de la compansión comparada con la técnica más simple de atenuar en forma lineal la señal en la entrada (para mantenerla en el intervalo lineal del canal) y amplificarla en forma lineal a la salida.

4.3 TRANSMISIÓN ANALÓGICA

Tratando todo junto, como hasta ahora, esta sección aplica los resultados de las anteriores al caso de la transmisión analógica. Por *transmisión analógica* se debe entender aquellos sistemas en los cuales las formas de onda que conducen la información se reproducen en el destino sin el empleo de técnicas de codificación digitales.

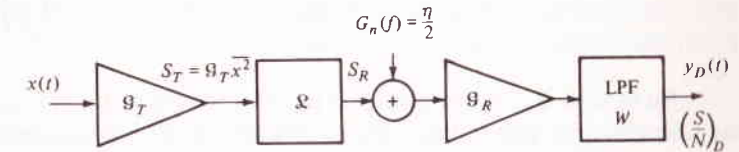



FIGURA 4.14
Sistema de transmisión analógica.

El análisis será en términos de una forma de onda arbitraria o *mensaje* designado por $x(t)$. Mejor aún, $x(t)$ representa el conjunto de mensajes probables de una fuente dada. Aunque tales mensajes no son estrictamente de banda limitada, es correcto suponer que existe alguna frecuencia superior —llámesele W— arriba de la cual el contenido espectral se pasa por alto, y no es necesario para conducir la información en cuestión. Así, se define

$$W = \text{ancho de banda de mensaje analógico}$$

en el sentido de que

$$G_x(f) \approx 0 \quad \text{para } |f| > W \tag{1}$$

Se supone además ergodicidad de tal manera que $\langle x^2(t)\rangle$ y $\overline{x^2}$ son intercambiables.

Relación señal a ruido

La figura 4.14 amplifica la 4.1 para la situación en cuestión. De manera específica, el transmisor resulta ser simplemente un amplificador con ganancia de potencia $\mathscr{G}_T$, tal que $S_T = \mathscr{G}_T \overline{x^2}$, y el filtro del receptor será un filtro pasabajos cercano al ideal con ancho de banda W, tal que $B_N \approx W$. Los otros parámetros son los mismos que se definieron en la sección 4.1.

Si el retardo de transmisión total es t_d y no hay distorsión sobre $|f| \leq W$, la señal de salida es

$$y_D(t) = \left(\frac{\mathscr{G}_T \mathscr{G}_R}{\mathscr{L}}\right)^{1/2} x(t - t_d) + n_D(t) \qquad (2)$$

de tal manera que las potencias de señal y ruido en el destino son

$$S_D = \frac{\mathscr{G}_T \mathscr{G}_R}{\mathscr{L}} \mathrm{E}[x^2(t - t_d)] = \frac{\mathscr{G}_T \mathscr{G}_R}{\mathscr{L}} \overline{x^2}$$

$$N_D = \overline{n_D{}^2} = \mathscr{G}_R \eta W$$

Por lo tanto

$$\left(\frac{S}{N}\right)_D = \frac{\mathscr{G}_T \overline{x^2}}{\mathscr{L}\eta W} = \frac{S_T}{\mathscr{L}\eta W} \qquad (3)$$

o, puesto que $S_T/\mathscr{L} = S_R$,

$$\left(\frac{S}{N}\right)_D = \frac{S_R}{\eta W} \qquad (4)$$

En la tabla 4.2 se listan valores representativos de $(S/N)_D$ para señales analógicas seleccionadas con el intervalo de frecuencia. El límite superior del intervalo de frecuencia es el valor nominal de W. El límite inferior también tiene significado en el diseño puesto que, por razón de los transformadores y capacitores de acoplamiento, muchos sistemas de transmisión analógica no responden conforme el sistema se acerca a la CD.

La ecuación (4) expresa $(S/N)_D$ en términos de algunos parámetros muy básicos, o sean, la potencia de la señal y la densidad del ruido a la entrada del receptor, y el ancho de banda del mensaje. Esta combinación en términos ocurrirá una y otra vez, en forma particular cuando se comparen varios tipos de sistemas; así, se le da un símbolo particular para su definición

$$\gamma \triangleq \frac{S_R}{\eta W} \qquad (5)$$

Aquí, γ es igual a $(S/N)_D$ para transmisión de banda base analógica. Se puede interpretar también el denominador ηW como la *potencia de ruido en el ancho de banda del mensaje*, aunque N_D difiere de ηW en $\mathscr{G}_R$. (Recuérdese que este factor de ganancia se cancela en las relaciones señal a ruido.)

Como la ecuación (4) presupone condiciones de transmisión sin distorsión, ruido blanco aditivo, y un filtro cercano al ideal, es más exacto decir que

$$\left(\frac{S}{N}\right)_D \leq \gamma \qquad (6)$$

En otras palabras, γ por lo general es un *límite superior* para la banda base analógica que puede o no alcanzarse en un sistema real. Por ejemplo, el ancho de banda de ruido en un filtro pasabajas práctico será algo mayor que el ancho de banda del mensaje, dando $(S/N)_D = S_R/\eta B_N < \gamma$. De manera similar, las no linealidades que causan que la salida incluya

Tabla 4.2 REQUISITOS DE TRANSMISIÓN TÍPICOS PARA SEÑALES ANALÓGICAS SELECCIONADAS

Tipo de señal	Intervalo de frecuencia	Relación señal a ruido, dB
Voz apenas inteligible	500 Hz–2 kHz	5–10
Voz en telefonía de buena calidad	200 Hz–3.2 kHz	25–35
Audio para radiodifusión de AM de buena calidad	100 Hz–5 kHz	40–50
Audio de alta fidelidad	20 Hz–20 kHz	55–65
Video en televisión	60 Hz–4.2 MHz	45–55

los términos veces-señal-ruido también reducen la relación señal a ruido efectiva. Por otra parte, la compansión puede dar lugar a mejoras de consideración†. Los efectos de la distorsión lineal y de ruido no blanco se examinan simultáneamente, a continuación.

Filtros terminales óptimos ★

Cuando el ruido es no blanco y/o el canal requiere de una considerable igualación, la aproximación más que sencilla hecha antes, debe sustituirse por una técnica más compleja en la cual los filtros diseñados en forma especial se incorporan a las dos terminales, el transmisor y el receptor. La figura 4.15 es el diagrama del sistema, con las ganancias de potencia $\mathscr{G}_T$, $1/\mathscr{L}$, y $\mathscr{G}_R$ esta vez absorbidas en las funciones de respuesta de frecuencia $H_T(f)$, $H_c(f)$ y $H_R(f)$.

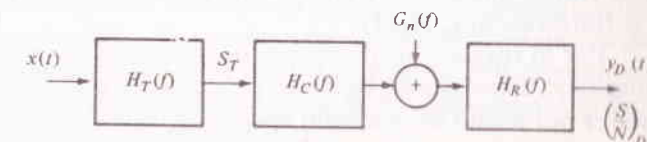



FIGURA 4.15

† Véase Bennett (1970, Cap. 3).

no habiendo razón para creer que $G_x(f)$ sea mayor o menor a una frecuencia particular. Prosiguiendo con esta suposición, se obtendría un buen diseño, pero no necesariamente el óptimo.

En realidad, la ecuación (15) es la suposición fundamental de la aproximación previa que se tuvo, donde se consideró al ruido como blanco y se dijo que el filtro receptor sirve sólo para eliminar el ruido fuera de banda. Aclarando este punto, sea $G_n(f) = \eta/2$ y sea el canal sin distorsión tal que $|H_c(f)|^2 = 1/\mathscr{L}$. Entonces, insertando en la ecuación (13) se demuestra que los filtros terminales resultan ser filtros pasabajas ideales en tanto el denominador de la ecuación 14 es

$$\left| \int_{-W}^{W} (\mathscr{L}\overline{x^2}\eta/4W)^{1/2}\, df \right|^2 = \mathscr{L}\overline{x^2}\eta W$$

tal que $(S/N)_{D_{\text{máx}}} = S_T/\mathscr{L}\eta W = S_R/\eta W = \gamma$.

EJERCICIO 4.3 Considérese un sistema que tiene $|H_c(f)|^2 = 1/\mathscr{L}$ y $G_n(f) = (\eta/2)(1 + a^2 f^2)^2$. Tomando $G_x(f)$ de la ecuación (15), encuéntrese $(S/N)_{D\text{máx}}$ en términos de S_R y compárese con $(S/N)_D$ cuando el filtro receptor es un filtro pasabajas ideal con $B = W$. *Respuesta*: $(S/N)_{D\text{máx}} = S_R/\eta W[1 + (a^2W^2/3)]^2$, $(S/N)_D = S_R/\eta W[1 + (2a^2W^2/3) + (a^4W^4/5)]$.

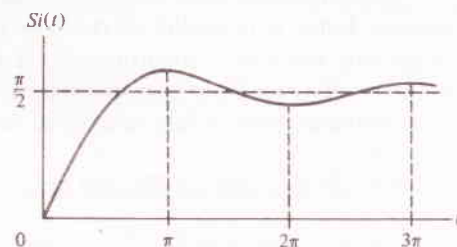



FIGURA 4.16
La integral seno, Si $(t) = \int_0^t (\text{sen}\,\lambda)/\lambda d\lambda$.

4.4 TRANSMISIÓN DE PULSOS

La transmisión de pulsos difiere de la transmisión analógica en que no interesa la reproducción fiel de la forma de onda del mensaje, sino más bien detectar la presencia de un pulso para analizar a dos o más pulsos espaciados en forma cercana, y medir su amplitud o su posición en el tiempo. Ejemplos de esto, son los sistemas telegráficos o de radar. Esta sección está dedicada a examinar los efectos del ancho de banda limitado y del ruido aditivo en transmisión de pulsos en banda base.

Requisitos de ancho de banda

Los pulsos cortos tienen grandes anchos espectrales, como se ha dicho una y otra vez. Invirtiendo esta observación, se puede decir que dado un sistema de ancho de banda fijo, existe un límite inferior en la duración de los pulsos a la salida, es decir una *duración de pulso de salida mínima*. En consecuencia, el número máximo de pulsos de salida distintos que se pueden analizar por unidad de tiempo está limitado por el ancho de banda del sistema.

Para situar el tema con un fundamento cuantitativo, sea $x(t) = A\Pi(t/\tau)$, un pulso rectangular, la entrada a un filtro pasabajas ideal o cercano al ideal con ancho de banda B, con ganancia unitaria y retardo en tiempo cero, tal que $H(f) = \Pi(f/2B)$. Puesto que el espectro de entrada $X(f) = A\tau\,\text{senc}\,f\tau$ tiene simetría par, la transformada de Fourier inversa para la salida $y(t) = \mathscr{F}^{-1}[H(f)X(f)]$ se simplifica a

$$y(t) = 2\int_0^B A\tau \frac{\text{sen}\,\pi f\tau}{\pi f\tau} \cos 2\pi ft\, df$$
$$= \frac{A}{\pi}\left[\int_0^B \frac{\text{sen}\,\pi f(2t+\tau)}{f} df - \int_0^B \frac{\text{sen}\,\pi f(2t-\tau)}{f} df\right]$$

la cual es aún una integral no elemental que requiere de evaluación en serie. Por fortuna, el resultado se puede expresar en términos del *seno integral* tabulado.

$$\text{Si}\,(t) \triangleq \int_0^t \frac{\text{sen}\,\lambda}{\lambda} d\lambda \qquad (1$$

según la gráfica de la figura 4.16. Cambiando las variables de integración se llega finalmente a

$$y(t) = \frac{A}{\pi}\{\text{Si}\,[\pi B(2t+\tau)] - \text{Si}\,[\pi B(2t-\tau)]\} \qquad (2$$

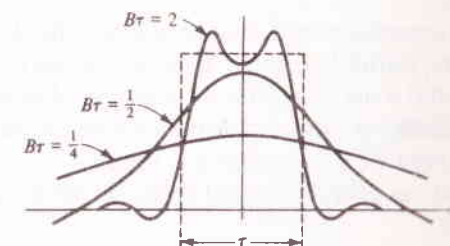



FIGURA 4.17
Respuesta al pulso de un filtro pasabajas ideal.

lo cual se muestra en la figura 4.17 para tres valores del producto $B\tau$. Nótense los precursores originados por el filtro ideal.

A despecho de involucrar más matemáticas, las conclusiones obtenidas de la figura 4.17 son muy simples. Se ha dicho que el ancho espectral de un pulso rectangular es en forma aproximada $1/\tau$. Para $B \gg 1/\tau$, la señal de salida es de manera esencial no distorsionada; mientras que para $B \ll 1/\tau$, el pulso de salida se extiende y tiene una duración que depende más del ancho de banda del filtro que de la señal de entrada. Como una regla a la mano, burda pero muy útil, se puede decir que la duración de pulso de salida mínima y el ancho de banda se relacionan por medio de

$$\tau_{\min} \geq \frac{1}{2B} \tag{3}$$

con tal que el pulso de entrada tenga $\tau \leq \tau_{\min}$. Yendo algo más lejos, se puede decir también que el número máximo de pulsos de salida analizados por unidad de tiempo es en forma aproximada $1/\tau_{\min} = 2B$. Esto se logra empleando pulsos de entrada de una duración menor que $\frac{1}{2}B$ y espaciados en tiempo en $\frac{1}{2}B$. La figura 4.18, que muestra las señales de entrada y salida para dos pulsos espaciados por τ, resiste esta afirmación.

Con la detección y análisis de pulsos, se puede tener interés en el problema de la ubicación de pulsos o medición de posición con respecto a algún tiempo de referencia. Por lo general, las mediciones de posición están basadas en el borde principal del pulso y para ese propósito son deseables los pulsos rectangulares, puesto que el borde tiene una posición única. Pero las formas de pulso realizables se elevan en forma más gradual hacia sus valores pico, ocasionando que la posición del borde principal sea ambigua y su medición menos evidente. La

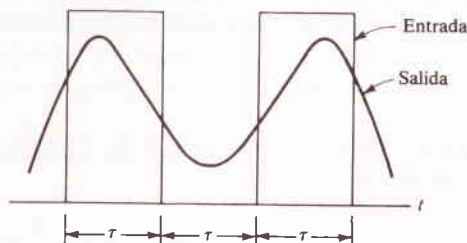



FIGURA 4.18
Análisis al pulso de un filtro pasabajas ideal, $B = \frac{1}{2}\tau$.

regla convencional acerca de la incertidumbre se establece en términos del *tiempo de subida*, definido como el intervalo requerido por el pulso para ir de cero a la amplitud total o del 10 al 90% de la amplitud total. Se dice entonces que la incertidumbre en cuanto a la medición de posición de pulsos es en forma aproximada igual al tiempo de subida τ_r. En relación a la figura 4.17 se ve que el tiempo de subida de un pulso filtrado es proporcional al ancho de banda. Por lo que, se tiene como otra regla a la mano,

$$\tau_{r\min} \geq \frac{1}{2B} \tag{4}$$

Cuando los pulsos de entrada tienen $\tau_r \leq \tau_{r\min}$, los pulsos de salida tendrán tiempos de subida no menores que $\frac{1}{2}B$ y la incertidumbre de ubicación mínima es aproximadamente $\frac{1}{2}B$. En forma alterna, si el tiempo de subida de la entrada es mayor que $\frac{1}{2}B$, el tiempo de subida de la salida será, aproximadamente, el mismo que el de entrada.

Queda convenido que las ecuaciones (3) y (4) son guías burdas basadas en el caso de una entrada rectangular a un filtro pasabajas ideal; las cuales son, no obs-

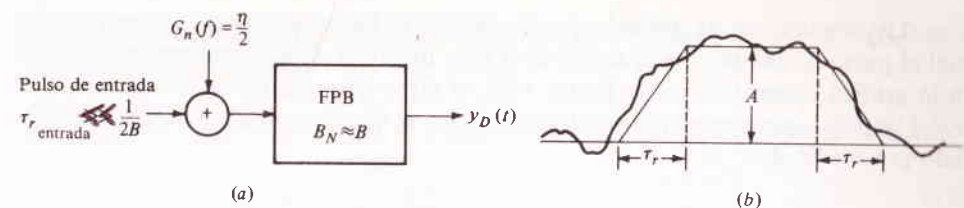



FIGURA 4.19
Mediciones de pulsos en ruido aditivo. (*a*) Diagrama a bloques; (*b*) pulso filtrado con ruido.

tante y en general, útiles. En estudios de otras formas de pulsos y otros filtros pasabajas, se demuestra que estas desigualdades son apropiadas, tomando B como el ancho de banda a 3 dB.

Mediciones de pulsos en ruido aditivo

Sea un pulso (no es necesario que sea rectangular) contaminado con ruido blanco aditivo y pasado a través de un fitro pasabajas cuyo ancho de banda equivalente de ruido y ancho de banda a 3 dB son casi iguales, figura 4.19*a*. Si el tiempo de subida del pulso de entrada es pequeño en comparación con $1/2B$, la forma del pulso de salida puede ser aproximada a un *trapezoide* más ruido $n_D(t)$, figura 4.19*b*. Así, las mediciones de la amplitud del pulso y de la posición serán erróneas obedeciendo al ruido.

En el pico del pulso de salida, $y_D(t) = A + n_D(t)$, se puede definir la amplitud cuadrática media normalizada de error como

$$\epsilon_A{}^2 \triangleq \frac{\overline{n_D{}^2}}{A^2} = \frac{\eta B}{A^2} \tag{5a}$$

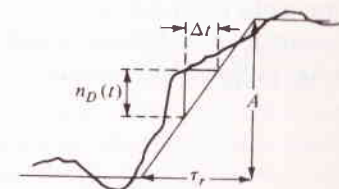



FIGURA 4.20
Vista ampliada de una perturbación de ruido.

Pero la energía del pulso de salida es $E \approx A^2\tau$ y $\tau \geq 1/2B$; en consecuencia

$$\epsilon_A{}^2 \geq \frac{\eta}{2E} \tag{5b}$$

lo cual da un límite inferior en el error. Este límite inferior es válido cuando $\tau = 1/2B$.

Las mediciones de posición por lo general se hacen anotando el tiempo en el cual el pulso de salida excede algún nivel fijo, digamos $A/2$. Entonces, como se ve en la gráfica aumentada de la figura 4.20, el error de posición en tiempo es $\Delta t = (\tau_r/A)n_D(t)$ —por semejanza de triángulos— y el error cuadrático medio normalizado por medio de τ^2 es

$$\epsilon_t^2 \triangleq \frac{\overline{\Delta t^2}}{\tau^2} = \overline{n_D^2}\frac{\tau_r^2}{A^2\tau^2} = \eta B\frac{\tau_r^2}{A^2\tau^2} \tag{6a}$$

Por último, insertando $\tau_r \geq 1/2B$ y $E = A^2\tau$ se llega al límite inferior

$$\epsilon_t^2 \geq \frac{\eta}{4BE\tau} \tag{6b}$$

En contraste con la ecuación (5*b*), este límite inferior se logra cuando $\tau_r = ½ B$ y $\tau \gg \tau_r$. Así, se debe emplear el ancho de banda más grande posible para reducir al mínimo los errores en la medición de posición.

Por lo pronto, el lector debe haberse percatado de que el análisis de la transmisión de pulsos es en gran manera intuitivo y de tipo exploratorio. Son posibles las investigaciones más refinadas de casos concretos, de manera notable el caso de la detección óptima de pulsos que se examina a continuación. Sin embargo, los resultados que se admitieron en su forma burda anteriormente, aún son valiosas guías para el diseño de sistemas de transmisión de pulsos, y las recíprocas de las ecuaciones (5) y (6) llevan en forma paralela el concepto de relación de señal a ruido de la transmisión analógica.

Detección óptima de pulsos-filtros acoplados ★

En forma similar a los filtros terminales óptimos para transmisión analógica, existe un filtro receptor óptimo para la detección de un pulso de forma conocida $x(t)$ contaminado con ruido aditivo con densidad espectral conocida $G_n(f)$. Tales filtros se designan como *filtros acoplados*, y se emplean en forma extensa en sistemas de radar y de transmisión de datos.

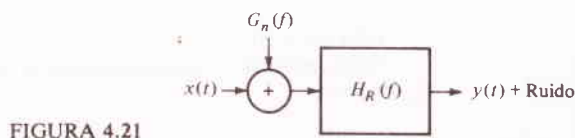



FIGURA 4.21

Considérese la situación de la figura 4.21, donde la forma del pulso de salida no es importante a menos que se desee aumentar su amplitud en algún tiempo arbitrario, digamos t_0, y reducir al mínimo el ruido de salida. En ausencia de ruido, la señal de salida pico en $t = t_0$ es

relaciona la ecuación (11) al estudio anterior y demuestra por qué a $H_{\text{ópt}}(f)$ se le llama un filtro acoplado.

EJERCICIO 4.4 Para el caso de ruido blanco $G_n(f) = \eta/2$, demuéstrese que

$$\left[\frac{|y(t_0)|^2}{N}\right]_{\text{máx}} = \frac{2E_x}{\eta} \tag{12a}$$

donde E_x es la energía en $x(t)$. También demuéstrese que

$$h_{\text{ópt}}(t) = \mathscr{F}^{-1}[H_{\text{ópt}}(f)] = Kx(t_0 - t) \tag{12b}$$

así, la respuesta de impulso tiene la misma forma que el pulso de entrada invertido en tiempo y desplazado en t_0. (*Sugerencia*: Empléese el hecho de que $v(-t) \longleftrightarrow V^*(f)$ cuando $v(t)$ es real.)

4.5 TRANSMISIÓN DIGITAL

Se concluye este capítulo con un breve examen de la transmisión digital en banda base. De manera fundamental, un mensaje digital no es otra cosa que una secuencia ordenada de *símbolos* obtenidos de un *alfabeto* de tamaño finito μ. (Por ejemplo, una fuente binaria tiene $\mu = 2$ y los símbolos alfabéticos son los dígitos **0** y **1**.) El objetivo de un sistema de comunicación digital es transmitir el mensaje en un lapso preestablecido con un mínimo de errores. Así, la *velocidad de señalización* y la *probabilidad de error* juegan el mismo papel en la transmisión digital, que el ancho de banda y la relación señal a ruido juegan en la transmisión analógica. Más aún, existe una estrecha relación entre la velocidad de señalización y el ancho de banda y entre la probabilidad de error y la relación señal a ruido.

Formas de onda y rapidez de señalización

Es normal pensar de una señal digital como en algo que es una hilera de pulsos rectangulares de amplitudes discretas. Y, en efecto, esa es a menudo la forma en que viene de una fuente de datos. A modo de ilustración, la figura 4.22*a* muestra el mensaje binario **10110100** como podría aparecer a la salida de un computador digital. Esta forma de onda, una secuencia simple de conectar-desconectar (*on-off*), se dice que es *unipolar*, porque tiene sólo una polaridad, y *sincrónico*, porque todos los pulsos tienen igual duración y no existe separación entre ellos. Las señales unipolares contienen una componente de CD diferente de cero que es difícil de transmitir, no conduce información y es un gasto inútil de potencia. De manera similar, las señales sincrónicas requieren coordinación en tiempo, en transmisor, y en receptor, lo cual significa complicaciones de diseño. La señal *polar* (dos polaridades) *con retorno a cero* de la figura 4.22*b* abarca ambos problemas, pero los "espacios" que efectúan la autosincronía de la señal son un desperdicio de tiempo de transmisión. Si la condición más importante a considerar es la eficiencia, será preferible la señal *sincrónica*

$$y(t_0) = \mathscr{F}^{-1}[H(f)X(f)]|_{t=t_0}$$

$$= \int_{-\infty}^{\infty} H(f)X(f)e^{+j\omega t_0}\, df \qquad (7)$$

donde $X(f) = \mathscr{F}[x(t)]$. (Se emplean las transformadas de Fourier más que la densidad espectral, dado que el pulso es una señal de energía conocida.) La potencia de ruido de salida es

$$N = \int_{-\infty}^{\infty} |H(f)|^2 G_n(f)\, df \qquad (8)$$

y la cantidad máxima es

$$\frac{|y(t_0)|^2}{N} = \frac{|\int_{-\infty}^{\infty} HXe^{j\omega t_0}\, df|^2}{\int_{-\infty}^{\infty} |H|^2 G_n\, df} \qquad (9)$$

donde $H(f)$ es la única función en el arreglo.

Para determinar $H_{\text{ópt}}(f)$, se recurre de nuevo a la desigualdad de Schwarz, ecuación (12), sección 4.3, esta vez en la forma

$$\frac{|\int_{-\infty}^{\infty} VW^*\, df|^2}{\int_{-\infty}^{\infty} |V|^2\, df} \leq \int_{-\infty}^{\infty} |W|^2\, df$$

cuyo primer miembro es el mismo que el de la ecuación (9) con

$$V = HG_n^{1/2} \qquad W^* = \frac{HXe^{j\omega t_0}}{V} = \frac{Xe^{j\omega t_0}}{G_n^{1/2}}$$

Dado que la desigualdad resulta ser una igualdad cuando $V(f) = KW(f)$, la relación en la ecuación (9) será máxima si

$$H_{\text{ópt}}(f) = K\frac{X^*(f)e^{-j\omega t_0}}{G_n(f)} \qquad (10)$$

donde K es una constante arbitraria, como lo es t_0. Por lo que,

$$\left[\frac{|y(t_0)|^2}{N}\right]_{\text{máx}} = \int_{-\infty}^{\infty} \frac{|X(f)|^2}{G_n(f)}\, df \qquad (11)$$

si el filtro se optimiza.

Obsérvese en la ecuación (10) que $H_{\text{ópt}}(f)$ da importancia a las frecuencias donde $|X(f)|/G_n(f)$ es grande, y viceversa, en una forma similar al filtro receptor óptimo para transmisión analógica, ecuación (13*a*), sección 4.3. De manera infortunada, $H_{\text{ópt}}(f)$ vuelve a ser menudo *irrealizable* † en forma física porque la correspondiente respuesta de impulso es diferente de cero para $t < 0$. El siguiente ejercicio

† Thomas (1969, Cap. 5) investiga el problema de la optimización con una limitación de realización.

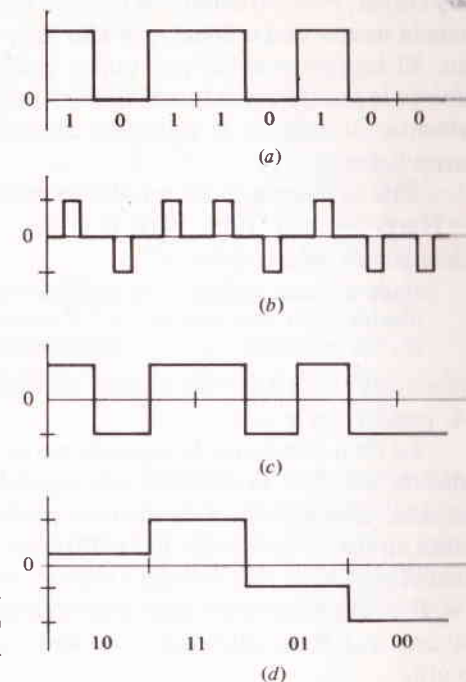



FIGURA 4.22
Formas de onda digitales. (*a*) Sincrónica unipolar; (*b*) polar con retorno a cero; (*c*) sincrónica polar; (*d*) cuaternaria sincrónica polar.

polar (Fig. 4.22*c*). Un caso ilustrado de nivel múltiple, lo vemos en la figura 4.22*d* en la cual una señal cuaternaria ($\mu = 4$) se deriva por agrupamiento de dígitos binarios en bloques de dos.

Haciendo caso omiso de los detalles específicos, la señal de entrada del canal es una representación analógica del mensaje digital que generalmente se puede describir como un tren de pulsos de la forma†

$$x(t) = \sum_k a_k p\left(t - \frac{k}{r}\right) \qquad (1)$$

donde a_k es el nivel de amplitud que representa el k-ésimo dígito del mensaje, $p(t)$ es la forma básica del pulso con valor pico $p(0) = 1$, el espaciamiento entre pulsos es $1/r$, y r es la velocidad de señalización. Por ejemplo, la señal sincrónica polar de la figura 4.22*c* tiene $a_k = \pm a$ y $p(t) = \Pi(t/\tau)$, donde la duración del pulso es $\tau = 1/r$. Una señal con retorno a cero tendría $\tau < 1/r$.

Si el canal es lineal y sin distorsión a todas las frecuencias; es decir, tiene ancho de banda infinito, entonces $p(t)$ no sufre degradación en la transmisión, y se puede lograr una velocidad de señalización arbitrariamente grande por el uso de pulsos

† El índice k indica la secuencia de tiempo, y sus límites, omitidos en la ecuación (1), que dependen del momento en que el mensaje empieza y termina.

muy cortos. Pero un canal real tiene un ancho de banda finito y una respuesta de frecuencia menor que la ideal, por ello los pulsos se dispersan o extienden y se traslapan. El ingeniero debe, por tanto, conformar la señal de salida para *reducir al mínimo la interferencia intersimbólica* debida al traslapamiento y, al mismo tiempo, *aumentar al máximo la velocidad de señalización*, objetivos que son mutuamente contradictorios.

Este problema se ha estudiado desde los primeros días de la telegrafía, pero fue Harry Nyquist (1924, 1928) el primero en establecer la relación ancho de banda-velocidad de señalización:

> Dado un canal pasabajas de ancho de banda B, es posible transmitir símbolos independientes a una velocidad de $r \leq 2B$ símbolos por segundo sin interferencia intersimbólica. No es posible transmitir símbolos independientes a $r > 2B$.

Nótese que $r \leq 2B$ está de acuerdo con la regla de resolución de pulsos de la sección 4.4, puesto que $\tau \leq 1/r$.

Es fácil demostrar la segunda parte de la relación, suponiendo que se trata de señalizar a $2(B + \epsilon)$ símbolos por segundo, siendo ϵ positiva pero arbitrariamente pequeña. Una secuencia de mensaje posible, está formada por dos símbolos alternantes en forma indefinida **01010101**, por ejemplo. La forma de onda resultante del canal es *periódica* con periodo $1/(B + \epsilon)$ y contiene sólo la frecuencia fundamental $f_0 = B + \epsilon$ más sus armónicas. Puesto que el canal no permite el paso de frecuencias mayores que B, la salida del canal será cero y una componente de CD posible pero no útil.

Para señalizar a la velocidad máxima $r = 2B$ se requiere una forma de pulso muy especial, o sea, el *pulso senc*

$$p(t) = \text{senc}\, rt \tag{2}$$

el cual es de *banda limitada* en $B = r/2$ y por lo mismo no sufre distorsión cuando se transmite. Por supuesto que $p(t)$ no es limitada en tiempo, pero tiene *cruzamientos periódicos cero*, es decir,

$$p\left(\frac{m}{r}\right) = \text{senc}\, m = \begin{cases} 1 & m = 0 \\ 0 & m \neq 0 \end{cases}$$

Así, si se forma la señal

$$x(t) = \sum_k a_k \,\text{senc}\, r\left(t - \frac{k}{r}\right) = \sum_k a_k \,\text{senc}\,(rt - k) \tag{3a}$$

entonces en cualquier tiempo $t = m/r$,

$$x\left(\frac{m}{r}\right) = \sum_k a_k \,\text{senc}\,(m - k) = a_m \tag{3b}$$

como se ilustra en la figura 4.23. En otras palabras, dados los cruzamientos cero, el traslapamiento de pulsos no ocasiona interferencia intersimbólica si se muestrea a $x(t)$ en forma periódica a la velocidad $r = 2B$.

Como puede inferir el lector, en este caso se necesita la información de tiempo entre transmisor y receptor; es decir, la señalización debe estar exactamente sincro-

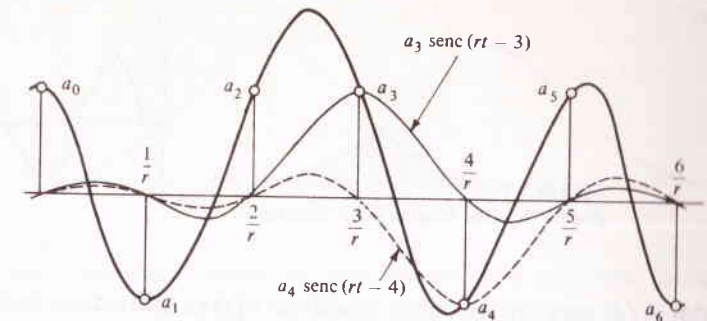



FIGURA 4.23
La forma de onda digital $x(t) = \sum_k a_k \,\text{senc}\,(rt - k)$.

nizada en $r = 2B$. Además, esta aproximación sólo es válida con un canal pasabajas ideal. El problema general de la formación de pulsos se trata adelante (Cap. 10), por ahora se tomará a $x(t)$ como en la ecuación (3), la cual viene a ser la elección más efectiva si el ancho de banda es limitado. Por otra parte, si el ancho de banda disponible resultara grande comparado con r, entonces los pulsos rectangulares serían la elección más conveniente.

EJERCICIO 4.5 Demuéstrese que la señalización sincrónica en $r = B$, es posible si $p(f) = \text{senc}^2\, rt$.

Ruido y errores

La figura 4.24 muestra los elementos básicos de un receptor de banda base digital. La señal recibida $K_R x(t)$ está contaminada por ruido aditivo pero un filtro pasabajas cercano al ideal elimina el ruido fuera de banda dando lugar a

$$y(t) = K_R\, x(t) + n(t)$$

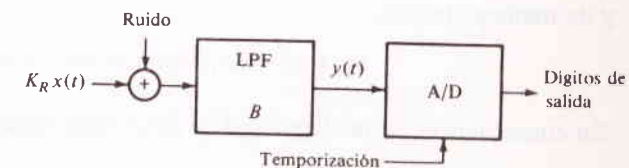



FIGURA 4.24
Receptor digital de banda base.

donde se han omitido los índices D por conveniencia. La señal analógica $y(t)$ se opera entonces por medio de un convertidor analógico a digital (A/D) cuya función es recuperar o *regenerar* el mensaje digital. Se proporciona sincronización al converti-

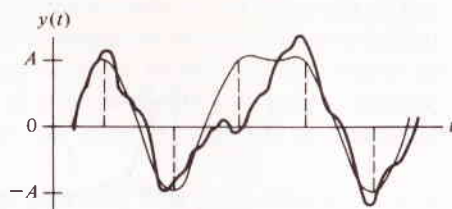



FIGURA 4.25
Forma de onda binaria polar más ruido.

dor A/D para que así pueda muestrear $y(t)$ en los tiempos óptimos $t = m/r$ cuando no haya interferencia intersimbólica, es decir, $y(m/r) = K_r a_m + n(m/r)$.

Para comenzar con un caso sencillo, considérese la recepción de una señal binaria polar con $K_R a_k = \pm A$ que representa los dígitos binarios **1** y **0**. En la figura 4.25 se ilustra una forma de onda típica de señal más ruido $y(t)$ como aparecería a la entrada de un convertidor A/D, suponiendo $\bar{n} = 0$. Una técnica de conversión directa es decidir en los momentos apropiados, si $y(m/r)$ está más cerca a $+A$ (intenta presumiblemente ser un 1) o más cerca a $-A$ (intenta presumiblemente ser un 0). De manera intuitiva, la *regla de decisión* lógica viene a ser: escójase **1** si $y(m/r) > 0$, escójase **0** si $y(m/r) < 0$, y láncese al aire una moneda si $y(m/r) = 0$. (Este último caso, por raro, no recibirá mayor atención.) El convertidor puede, por lo tanto, tomar la forma de un circuito de decisión sincronizada cuyo cruce o *nivel de umbral* se pone en cero, y los errores de conversión ocurren siempre que el ruido venga a ser $y(m/r)$ en el lado erróneo del umbral en el tiempo de decisión.

Así, en lo que a probabilidades de error se refiere, se tienen *dos* variables aleatorias †

$$y_1 = A + n \qquad y_0 = -A + n \tag{4}$$

correspondientes a los dígitos intentados **1** y **0**. Entonces, si fue intentado un **1**, la *probabilidad condicional* de error de conversión es

$$P_{e_1} \triangleq P(\text{error} \,|\, \mathbf{1} \text{ sent}) = P(y_1 < 0) = P(A + n < 0) \tag{5a}$$

y de manera similar

$$P_{e_0} \triangleq P(\text{error} \,|\, \mathbf{0} \text{ sent}) = P(y_0 > 0) = P(-A + n > 0) \tag{5b}$$

En consecuencia, *la probabilidad de error neta* resulta ser

$$P_e = P_1 P_{e_1} + P_0 P_{e_0} \tag{6}$$

donde P_1 y P_0 son las probabilidades de dígitos en la fuente, no necesariamente iguales aunque generalmente lo son. En cualquier caso $P_1 + P_0 = 1$, puesto que uno u otro deben ser transmitidos.

† Aquí, por sencillez, se emplean letras minúsculas para simbolizar las variables aleatorias.

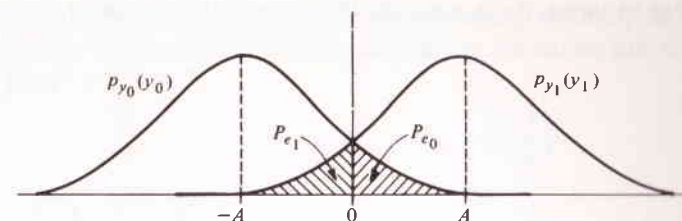



FIGURA 4.26
Funciones de densidad de probabilidad señal-más-ruido para señalización binaria polar.

En este punto se supone que $n(t)$ es un proceso *gaussiano* de valor medio cero con variancia σ^2, así su función de densidad de probabilidad es

$$p_n(n) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-n^2/2\sigma^2} \tag{7}$$

Esta suposición no es del todo falsa para sistemas lineales de banda base, puesto que la mayor parte de los ruidos eléctricos son gaussianos y las funciones gaussianas son invariantes bajo operaciones lineales. Se desprende de la ecuación (4) que y_1 y y_0 son también gaussianas con la misma variancia σ^2 pero con valores medios

$$\bar{y}_1 = +A \qquad \bar{y}_0 = -A$$

La figura 4.26 resume la situación de manera conveniente al mostrar las dos funciones de densidad de probabilidad, $p_{y1}(y_1) = p_n(y_1 - A)$ y $p_{y0}(y_0) = p_n(y_0 + A)$ de las cuales se calcularán P_{e_1} y P_{e_0}.

Recordando la interpretación de área de las funciones de densidad de probabilidad,

$$P_{e_0} = P(y_0 > 0) = \int_0^\infty p_{y_0}(y_0)\, dy_0$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma} \int_0^\infty e^{-(y_0 + A)^2/2\sigma^2}\, dy_0$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma} \int_{A/\sigma}^\infty e^{-\lambda^2/2}\, d\lambda$$

donde se ha efectuado el cambio de variable $\lambda = (y_0 + A)/j$. Esto pone a P_{e_0} en la misma forma que la función $Q(k)$, ecuación (8) (Secc. 3.4), con $k = A/\sigma$. Además, notando la simetría de la figura 4.26, $P(y_1 < 0) = P(y_0 > 0)$ tal que

$$P_{e_1} = P_{e_0} = Q\left(\frac{A}{\sigma}\right) \tag{8a}$$

Por lo tanto, de la ecuación (6), la probabilidad de error neta es

$$P_e = (P_1 + P_0)Q\left(\frac{A}{\sigma}\right) = Q\left(\frac{A}{\sigma}\right) \tag{8b}$$

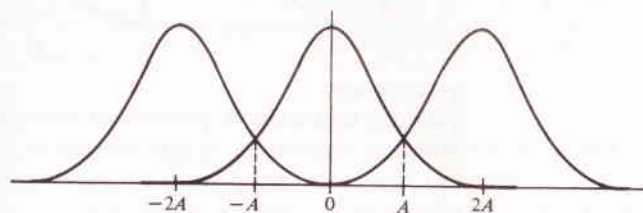



FIGURA 4.27
Funciones de densidad de probabilidad señal-más-ruido para señalización ternaria polar.

la cual es independiente de las probabilidades de dígito porque $P_{e_1} = P_{e_0}$. En este análisis hay varios puntos que merecen un comentario adicional:

1 Tomando el nivel de umbral en cero, se llega a *probabilidades de error iguales* para cada dígito, $P_{e_1} = P_{e_0}$.
2 Éste es el nivel de umbral *óptimo* si los dígitos son equiprobables ($P_1 = P_0$) puesto que cualquier otra elección incrementaría a P_{e_1} más de lo que disminuye P_{e_0}, o viceversa.
3 Las señales binarias en ruido gaussiano tienen una propiedad única en la figura 4.26. Si la amplitud intentada es $+A$, las excursiones de ruido positivas superpuestas no tienen efecto perjudicial; y de manera similar para $-A$ con excursiones de ruido negativas. En consecuencia, como $n(t)$ es igualmente probable que sea positiva o negativa, el convertidor corregirá por lo menos la mitad del tiempo, sin importar el ruido. Sin embargo, se debe recordar siempre que un mensaje binario con cincuenta por ciento de errores es ciento por ciento inútil.
4 En vista de la ecuación (8*b*) se pueden usar en forma directa la figura 3.5 o la tabla D como una gráfica de P_e contra A/σ y es evidente que P_e disminuye dramáticamente conforme aumenta A/σ. Si $A/\sigma = 2.0$, por ejemplo, $P_e \approx 2 \times 10^{-2}$ mientras que si $A/\sigma = 4.0$, $P_e \approx 3 \times 10^{-5}$. En forma incidental, muchas aplicaciones requieren probabilidades de error del orden de 10^{-4} o menores.

Dado que se ha tenido suficiente cuidado al examinar los errores de las señales binarias, la extensión de éstas a niveles múltiples o de orden μ, es bastante directa, siempre y cuando el ruido sea gaussiano. Considérese, por ejemplo, una señal ternaria polar ($\mu = 3$) con amplitudes de pulsos de salida $K_R a_k = +A$, 0, o $-2A$, que representan a los dígitos ternarios **2**, **1** y **0**, respectivamente. El equivalente a la figura 4.26 tiene tres funciones de densidad gaussiana (Fig. 4.27), y se ve que se necesitan

dos niveles de umbral. Bajo la condición usual de dígitos equiprobables, se encuentran con facilidad los niveles de umbral óptimos mostrados, los cuales son $\pm A$, para lo cual

$$P_{e_2} = P_{e_0} = Q\left(\frac{A}{\sigma}\right)$$

mientras que

$$P_{e_1} = 2Q\left(\frac{A}{\sigma}\right)$$

dado que ambas excursiones, tanto positivas como negativas, de ruido dan lugar a errores cuando $K_R a_k = 0$. (Por medio de otra elección de niveles de umbral es posible igualar las probabilidades de error por dígito, pero al precio de incrementar la probabilidad neta P_e.) Por lo que

$$P_e = P_2 P_{e_2} + P_1 P_{e_1} + P_0 P_{e_0} = \frac{4}{3}\,Q\left(\frac{A}{\sigma}\right)$$

donde se ha insertado $P_2 = P_1 = P_0 = 1/3$.

Generalizando a un μ arbitrario con probabilidades de dígitos iguales, un razonamiento similar da

$$P_e = 2\left(1 - \frac{1}{\mu}\right)Q\left(\frac{A}{\sigma}\right) \tag{9}$$

donde el *espaciamiento* entre amplitudes de pulsos de salida adyacentes es $2A$ —*así*, $2(\mu - 1)A$ es el valor pico a pico— y los $\mu - 1$ niveles de umbral estan centrados entre las amplitudes de pulsos. La ecuación (9) se reduce, claramente, a la (8*b*) cuando $\mu = 2$. Sin embargo, no queda dicho todo, por tres razones: primera, en general un dígito de orden μ representa más información que un dígito binario; segunda, existen diferentes severidades de error en sistemas de orden μ, dependiendo de los corrimientos del ruido de las amplitudes aparentes en uno o más pasos; tercera, no es explícita la relación entre la probabilidad de error y la relación señal a ruido. Una expresión analítica de las dos primeras es bastante difícil, pero la tercera se puede tratar como sigue.

Relaciones señal a ruido

Si el ruido contaminante es blanco, la potencia de ruido filtrado es

$$N = \eta B \tag{10a}$$

y

$$\sigma = \sqrt{N} = \sqrt{\eta B} \tag{10b}$$

puesto que se supuso $\overline{n} = 0$. Si el ruido no es blanco, el mejor filtro receptor no es un pasabajas ideal. El problema del filtrado *óptimo* para transmisión digital con espectro de ruido arbitrario se trata páginas adelante (Cap. 10).

Para calcular la potencia de la señal, se necesita hacer una breve confrontación de las propiedades de $p(t)$ = senc rt. Se demuestra con facilidad (Prob. 4.32) que

$$\int_{-\infty}^{\infty} A_k p\left(t-\frac{k}{r}\right) A_m p\left(t-\frac{m}{r}\right) dt = \begin{cases} \dfrac{A_k^2}{r} & m = k \\ 0 & m \neq k \end{cases} \tag{11}$$

lo cual significa que los pulsos que constituyen a $K_R x(t)$ son *mutuamente ortogonales* con energía A_k^2/r por pulso. Pero la ortogonalidad es una condición suficiente para la *superposición de energía*, y el mensaje de M dígitos de longitud, digamos

$$K_R x(t) = \sum_{k=1}^{M} A_k p\left(t - \frac{k}{r}\right)$$

tiene energía

$$E_M = \frac{1}{r} \sum_{k=1}^{M} A_k^2$$

lo cual también sería verdadero para pulsos *rectangulares* con $\tau = 1/r$. Entonces, puesto que la potencia promedio es igual a la energía por unidad de tiempo y un mensaje de M dígitos es de M/r segundos de longitud, la potencia de señal recibida es

$$S_R = \frac{E_M}{(M/r)} = \frac{1}{M} \sum_{k=1}^{M} A_k^2$$

o simplemente el *promedio* de A_k^2. Generalizando, se tiene que si

$$K_R x(t) = \sum_{k=-\infty}^{\infty} A_k \operatorname{senc}(rt - k) \tag{12a}$$

y si los A_k son estadísticamente independientes, entonces

$$S_R = K_R^2 \overline{x^2} = \overline{A^2} \tag{12b}$$

donde $\overline{A^2}$ es el *promedio* estadístico o *ensamble* de A_k^2. Conforme sea necesario, este resultado se empleará varias veces en capítulos subsecuentes.

En el caso de una onda binaria polar con $A_k = \pm A$, igualmente probable,

$$\overline{A^2} = (+A)^2 P(+A) + (-A)^2 P(-A) = \frac{A^2}{2} + \frac{A^2}{2} = A^2$$

En el caso de orden μ polar con μ par,

$$A_k = \pm A, \pm 3A, \ldots, \pm(\mu - 1)A$$

y suponiendo símbolos equiprobables como $P(A_k) = 1/\mu$, aplicando la ecuación (1) (Sec. 3.3), da

$$\overline{A^2} = \sum_{k=1}^{\mu} A_k^2 P(A_k) = \frac{2}{\mu} \sum_{k=1}^{\mu/2} (2k-1)^2 A^2$$

Así, con la ayuda de las sumas de las tablas B, se tiene

$$\overline{A^2} = \frac{\mu^2 - 1}{3} A^2 \tag{13}$$

La ecuación (13) también es válida para cualquier onda polar, μ par o impar, con espaciamiento de amplitud uniforme $2A$.

Por último, combinando las ecuaciones (10), (12) y (13), la relación señal a ruido en la entrada del convertidor A/D es

$$\frac{S}{N} = \frac{S_R}{\eta B} = \frac{\mu^2 - 1}{3} \left(\frac{A}{\sigma}\right)^2 \tag{14}$$

y la ecuación (9) se puede escribir como

$$P_e = 2\left(1 - \frac{1}{\mu}\right) Q\left(\sqrt{\frac{3}{\mu^2 - 1} \frac{S}{N}}\right) \tag{15}$$

En forma alternativa, es útil definir el parámetro del sistema

$$\rho \triangleq \frac{S_R}{\eta r} \tag{16}$$

donde r es la rapidez de señalización y, en consecuencia, S_R/r es igual a la energía recibida promedio por dígito. El parámetro ρ juega, en esencia, el mismo papel en la transmisión digital que en la analógica. Para la forma del pulso en cuestión, $r = 2B$ como $S_R/\eta B = 2S_R/\eta r = 2\rho$ y la ecuación (15) resulta ser

$$P_e = 2\left(1 - \frac{1}{\mu}\right) Q\left(\sqrt{\frac{6\rho}{\mu^2 - 1}}\right) \tag{17a}$$

$$= Q(\sqrt{2\rho}) \qquad \mu = 2 \tag{17b}$$

De nuevo, como en la transmisión analógica, la ecuación (15) o la (17) son un límite superior en el comportamiento del sistema —o sea, un *límite inferior* en P_e— y las varias imperfecciones ocasionarán que la probabilidad de error sea mayor que la predicha. En particular, la potencia de CD desperdiciada en una forma de onda *unipolar* cambia a la figura en forma apreciable.

EJERCICIO 4.6 Considérese un sistema binario unipolar con $A_k = 2A$, o 0, y el umbral de decisión en A. Demuéstrese que $S_R = 2A^2$ si $P_1 = P_0$, y

$$P_e = Q\left(\frac{A}{\sigma}\right) = Q\left(\sqrt{\frac{S}{2N}}\right) = Q(\sqrt{\rho}) \tag{18}$$

Después calcúlese P_e para un sistema polar y unipolar, teniendo ambos $\rho = 8.0$ *Respuesta*: 3×10^{-5}, 2×10^{-3}.

Ejemplo 4.2

Se desea transmitir dígitos cuaternarios ($\mu = 4$) a una rapidez de 5 000/s en un sistema que tiene $S_R/\eta = 80\,000$. Suponiendo que se utilizan pulsos senc, el ancho de banda necesario es $B = r/2 = 2\,500$, y aplicando las ecuaciones (16) y (17) se tiene

$$\rho = \frac{S_R}{\eta r} = 16.0$$

$$P_e = \frac{3}{2} Q\left(\sqrt{\frac{2\rho}{5}}\right) \approx 10^{-2}$$

por lo que el sistema no es muy confiable.

Supóngase, sin embargo, que cada dígito cuaternario se sustituye por *dos* dígitos binarios. La rapidez buscada es entonces $r' = 2 \times 5\,000 = 10\,000$ y

$$\rho' = \frac{S_R}{\eta r'} = 8.0$$

$$P_e' = Q(\sqrt{2\rho'}) \approx 3 \times 10^{-5}$$

lo cual es una mejora sustancial en el comportamiento. El precio de esta mejora es un *ancho de banda mayor*, $B' = 2B$, más las unidades de *codificación* y de *decodificación*.

Se generalizarán estas observaciones adelante, en los títulos *Reducción de ruido de banda ancha* y *Teoría de la información*.

Repetidores regenerativos

Para la transmisión a larga distancia se necesitan repetidores, tanto para la comunicación analógica como la digital. Pero a semejanza de los repetidores de mensajes analógicos, los digitales pueden ser *regenerativos* en el sentido de que cada uno tiene un convertidor A/D y un amplificador. Si la probabilidad de error por repetidor es razonablemente baja y el número de saltos M es grande, la ventaja de la regeneración resulta mejor. Esto se demostrará en el caso de la transmisión binaria polar.

Cuando se emplean repetidores analógicos y se aplica la ecuación (17), (Secc. 4.1), la relación señal a ruido final es

$$P_e = Q\left[\sqrt{\frac{1}{M}\left(\frac{S}{N}\right)_1}\right] \qquad (19)$$

donde $(S/N)_1$ es la relación señal a ruido después de un salto. Por lo tanto, la potencia transmitida *por repetidor* se debe incrementar linealmente con M de valor par, un factor que no se debe dejar de tomar en cuenta cuando, por ejemplo, se necesitan 100 o más repetidores para cruzar el continente. El término $1/M$ en la ecuación (19)

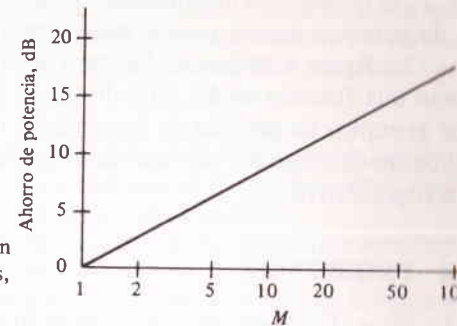



FIGURA 4.28
Ahorro de potencia ganado por regeneración como una función del número de repetidores, $P_e = 10^{-5}$.

procede del hecho de que el ruido contaminante se produce en forma progresiva de repetidor a repetidor.

En cambio, una estación repetidora regenerativa está compuesta de un receptor y un transmisor completos unidos en un paquete. La parte receptora convierte las señales entrantes a dígitos de mensaje, pero al mismo tiempo añade algunos errores al proceso; los dígitos se entregan después a la parte transmisora, la cual a su vez genera una nueva señal que transmite a la siguiente estación. Así, la señal regenerada queda completamente libre de ruido aleatorio pero contiene algunos errores.

Para analizar el comportamiento, sea ϵ la probabilidad de error en cada repetidor, o sea,

$$\epsilon = Q\left[\sqrt{\left(\frac{S}{N}\right)_1}\right] \qquad (20)$$

suponiendo unidades idénticas. Conforme pasa un dígito de estación a estación, el repetidor acumula errores de conversión. Si el número de conversiones erróneas es *par*, se cancelan, y se entrega un dígito correcto al destino. (Nótese que esto es sólo válido para datos *binarios*). La probabilidad de n errores en M conversiones sucesivas está dada por la *distribución binomial* de la ecuación (1), (Sec. 3.4):

$$P_M(n) = \binom{M}{n}\epsilon^n(1-\epsilon)^{M-n}$$

La probabilidad de error neta es entonces la probabilidad de que n sea *impar*; en forma específica,

$$P_e = \sum_{n \text{ impar}} P_M(n) = \binom{M}{1}\epsilon(1-\epsilon)^{M-1} + \binom{M}{3}\epsilon^3(1-\epsilon)^{M-3} + \cdots \approx M\epsilon$$

donde se aplica la aproximación para $\epsilon \ll 1$ y M no muy grande. En consecuencia, insertando la ecuación (20),

$$P_e \approx MQ\left[\sqrt{\left(\frac{S}{N}\right)_1}\right] \qquad (21)$$

así, P_e se incrementa linealmente con M, lo cual por lo general necesita un incremento de potencia mucho menor para contrarrestarla de la ecuación (19).

La figura 4.28 ilustra el ahorro de potencia proporcionado por la regeneración como una función de M, estando fija la probabilidad de error en $P_e = 10^{-5}$. Así, por ejemplo, un sistema de banda base no regenerativo de 10 estaciones, necesita aproximadamente 8.5 dB más de potencia transmitida (por repetidor) que un sistema regenerativo.

4.6 PROBLEMAS

4.1 (Secc. 4.1) Un sistema por cable de 20 km tiene S_t = 10 dBW, α = 0.64, y B_N = 500 kHz. El ruido es aditivo y blanco con $\mathcal{T}_N = 25\mathcal{T}_0$. Encuéntrense $(S/N)_D$ y $\mathcal{G}_P$ en decibeles tales que S_D = 0 dBW.
Respuesta: 33, 100

4.2 (Secc. 4.1) Repítase el problema 4.1 con α = 0.32 y B_N = 5 MHz.

4.3 (Secc. 4.1) Un sistema repetidor de 400 km emplea un cable con 8.68 α = 0.5. Si $S_T/\eta_1 B_N$ = 70 dB, ¿cuál es el número mínimo de repetidores tal que $(S/N)_D \geq 20$ dB? Supóngase un espaciamiento entre repetidores igual, o sea, $\mathcal{L}_1 = \mathcal{L}^{1/M}$.

4.4★ (Secc. 4.1) Estando fijos todos los otros parámetros, demuéstrese que el número de repetidores igualmente espaciados que aumenta al máximo a $(S/N)_D$ es $M = \ln\mathcal{L} = 0.23\ \mathcal{L}_{dB}$.

4.5★ (Secc. 4.1) Obedeciendo a no linealidades, la salida de un sistema de banda base es $y_D(t) = x(t) + n(t) + \frac{1}{2}[x(t) + n(t)]^2$. Encuéntrese $\overline{y_D^2}$ suponiendo que $x(t)$ y $n(t)$ son independientes y todos sus momentos impares son cero. ¿Es posible definir a $(S/N)_D$ en este caso?

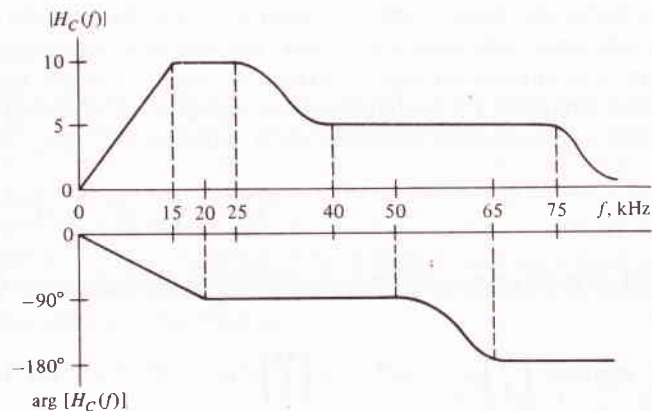



FIGURA P4.1

4.6★ (Secc. 4.1) La salida de señal a circuito abierto de un oscilador es $A \cos 2\pi f_0 t$. La resistencia de la fuente es R_s y existe ruido térmico generado internamente con temperatura $\mathcal{T}_N$. Se conecta un capacitor C a través de las terminales de salida para mejorar la S/N.
(*a*) Obténgase una expresión para S/N.
(*b*) ¿Qué valor de C aumenta al máximo a S/N.

4.7 (Secc. 4.2) La figura P4.1 muestra la relación de amplitudes y corrimiento de fase de un cierto canal de transmisión. ¿En qué intervalo o intervalos de frecuencia se tiene: distorsión de amplitud, distorsión de fase, transmisión sin distorsión?

4.8 (Secc. 4.2) Demuéstrese que un filtro pasabajas RC da en forma esencial transmisión sin distorsión si $x(t)$ es de banda limitada en $W \ll B = \frac{1}{2}\pi$RC.

4.9 (Secc. 4.2) La entrada a un filtro pasabajas RC es la señal de prueba de la figura 4.4 Constrúyase gráfica de la forma de onda de salida cuando $f_0 = B/3$.

4.10 (Secc. 4.2) Encuéntrese $t_d(f)$ para un filtro pasabajas RC y evalúesele en $f = B/4$, B, y $4B$ cuando B = 1 kHz.

4.11 (Secc. 4.2) Considérese una función de transferencia con *ondulaciones en la relación de amplitudes*, $H(f) = (1 + 2\,\alpha \cos \omega t_0) e^{-j\omega t_d}$, $|\alpha| \leq \frac{1}{2}$.
(*a*) Demuéstrese que $y(t) = x(t - t_d) + \alpha x(t - t_d + t_0) + \alpha x(t - t_d - t_0)$, tal que hay un par de "ecos".
(*b*) Tomando $\alpha = \frac{1}{2}$ y $x(t) = \Pi(t/\tau)$, bosquéjese $y(t)$ para $t_0 = 2\tau$, $\tau/2$, y $\tau/4$.

4.12★ (Secc. 4.2) Encuéntrese $y(t)$ en términos de $x(t)$ cuando hay *pequeñas ondulaciones en el corrimiento de fase*, es decir, $H(f) = \exp[-j(\omega t_d - \alpha \operatorname{sen} \omega t_0)]$, $|\alpha| \ll \pi$. Compárese con el problema 4.11. (*Sugerencia*: Empléese un desarrollo en serie para $(j\alpha \operatorname{sen} \omega t_0)$.)

4.13 (Secc. 4.2) Bosquéjese $|H_{eq}(f)|$ y $[H_{eq}(f)]$ necesarios para igualar a $H_c(f)$ en la figura P4.1 sobre $5 < |f| < 25$ kHz.

4.14 (Secc. 4.2) Supóngase que $x(t) = v(t) + \cos 2\pi f_1 t$ se aplica a un sistema no lineal con $y(t) = x(t) + 0.4x^2(t) + 0.1x^3(t)$.
(*a*) Encuéntrese $y(t)$ y bosquéjese una $y(f)$ típica cuando $v(t)$ es de banda limitada en $W \ll f_1$.
(*b*) Si $v(t) = \cos 2\pi f_2 t$, $f_2 > f_1$, enumérense todas las componentes de frecuencia en $y(t)$.

4.15 (Secc. 4.3) Un sistema diseñado para transmisión de voz con buena calidad telefónica tiene $(S/N)_D$ = 30 dB cuando $S_T = -3$ dBW. Si el ancho de banda se incrementa en forma apropiada, ¿qué valor de S_T se requiere para mejorar el sistema en cuanto a transmisión de audio de alta fidelidad, permaneciendo todos los otros factores sin cambio?
Respuesta: 30 a 40 dBW.

4.16 (Secc. 4.3) Un sistema diseñado para una señal analógica con W = 10 kHz emplea un filtro pasabajas RC con 3dB de ancho de banda B = 15 kHz en el receptor.
(*a*) Encuéntrese $(S/N)_D$ en términos de γ.
(*b*) Repítase para un filtro Butterworth de segundo orden (ejercicio 3.10, Secc. 3.6) con B = 12 kHz.

4.17 (Secc. 4.3) Un canal en que hay distorsión con $H_c(f) = [1 + j(2f/W)]^{-1}$ y ruido blanco se emplea para una señal con $G_x = (\frac{1}{2}W)\Pi(f/2W)$. Para compensar la distorsión, el filtro receptor es $H_R(f) = [1 + j(2f/W)\Pi(f/2W)$. El filtro transmisor es simplemente un amplificador con ganancia unitaria. Demuéstrese que $(S/N)_D \approx 0.78\gamma$.

4.18 (Secc. 4.3) Con relación a la figura 4.15, obténganse expresiones para $\mathcal{G}_T$, $\mathcal{L}$, y $\mathcal{G}_R$ en términos de $H_T(f)$, $H_c(f)$, $H_R(f)$ y $G_x(f)$. (*Sugerencia*: Por definición, $\mathcal{G}_T = S_T/\overline{x^2}$, etc.)

4.19 (Secc. 4.3) Dados $H_c(f)$ y $G_x(f)$ en el problema 4.17, encuéntrense los filtros terminales óptimos y evalúese $(S/N)_{D\,\text{máx}}$ en términos de γ.

4.20 (Secc. 4.4) Dibújese nuevamente la figura 4.18 para el caso de un filtro pasabajas RC con $B = \frac{1}{2}\tau$ y $B = \frac{1}{4}\tau$.

4.21 (Secc. 4.4) Se aplica un pulso senc $x(t) = A$ senc $2\,Wt$ a un filtro pasabajas ideal que tiene un ancho de banda B. Tomando la duración del pulso senc at como $\tau \triangleq 2/a$, constrúyase la gráfica de la relación de duración de salida a la de entrada como una función de B/W.

4.22 (Secc. 4.4) Un pulso rectangular con $\tau \gg 1/B$ se aplica a un filtro pasabajas RC. Encuéntrese del 10 al 90% del tiempo de subida en términos del ancho de banda a 3 dB.

4.23★ (Secc. 4.4) Algunas veces es conveniente para propósitos de análisis, definir el *ancho de banda efectivo* de un filtro pasabajas y la *duración efectiva* de su respuesta al impulso como sigue:

$$B_{ef} \triangleq \frac{\int_{-\infty}^{\infty}|H(f)|df}{2|H(f)|_{\text{máx}}} \qquad \tau_{ef} \triangleq \frac{\int_{-\infty}^{\infty} h(t)\,dt}{|h(t)|_{\text{máx}}}$$

Demuéstrese que $B_{ef}\tau_{ef} \geq \frac{1}{2}$. (*Sugerencia*: Primero demuéstrese que

$$\left|\int_{-\infty}^{\infty} H(f)\exp(j\omega t)\,df\right| \leq \int_{-\infty}^{\infty} |H(f)|\,df$$

4.24 (Secc. 4.4) La entrada a un filtro pasabajas RC es un pulso rectangular más ruido blanco.

(*a*) Demuéstrese que $\epsilon_A{}^2 = \pi B\tau(1 - e^{-2\pi B\tau})^{-2}$, donde E es la energía del *pulso de entrada*.

(*b*) Constrúyase gráfica de $\epsilon_A{}^2$ contra $B\tau$ y encuéntrese el mejor valor de $B\tau$ para mediciones de amplitud.

4.25 (Secc. 4.4) Se va a diseñar un sistema de transmisión de pulsos de manera que $E_A{}^2 \leq 10^{-2}$ y $\overline{\Delta t^2} \leq 10^{-12}$ cuando $\eta = 10^{-4}$. Suponiendo pulsos rectangulares en la entrada del filtro receptor, encuéntrense los valores mínimos para B, τ y E.

4.26 (Secc. 4.4) Se va a diseñar un filtro acoplado para el caso de $x(t) = \Pi[(t - \tau/2)/\tau]$ y $G_n(f) = \eta/2$. Bosquéjese $h_{\text{ópt}}(t)$ y determine la condición en t_0 tal que el filtro sea físicamente realizable, por lo menos en principio. Bosquéjese $y(t)$ en ausencia de ruido.

4.27 (Secc. 4.4) Supóngase que el sistema del problema 4.26 emplea un filtro pasabajas RC con $B = \frac{1}{2}\tau$ en lugar del filtro acoplado. Encuéntrese $|y(t_0)|^2/N$ en términos del máximo valor teórico posible.

4.28 (Secc. 4.5) Considérese una señalización digital con una forma de pulso gaussiano $p(t) = e^{-\pi(t/\tau)^2}$, el cual no es limitado en tiempo ni de banda limitada y no tiene *cruzamientos cero periódicos*. Se requiere que $p(m/r) \leq 0.01$ para reducir al mínimo la interferencia intersimbólica y se define el ancho de banda B tal que $|P(f)| \leq 0.01\,|P(0)|$ para $|f| \geq B$. Encuéntrese r en términos de B bajo estas condiciones.
Respuesta: $r = (\pi/2 \ln 10)\,B \approx 0.7B$.

4.29 (Secc. 4.5) Por medio de un bosquejo como el de la figura 4.26, demuéstrese gráficamente que cualquier nivel de umbral $b \neq 0$ da lugar a $P_e > Q(A/\sigma)$. ¿Puede usted demostrar esto en forma analítica?

4.30 (Secc. 4.5) Algunas veces los sistemas digitales presentan *ruido de impulso* cuya función de densidad de probabilidad se puede aproximar a $p(n) = (1/\sqrt{2}\sigma)\exp(-\sqrt{2}/n/\sigma)$. Encuéntrese P_e para un sistema binario polar con ruido de impulso, y compárese con $Q(A/\sigma)$.

4.31 (Secc. 4.5) Demuestre la ecuación (11) empleando la ecuación (13, Sec. 2.3).

4.32★ (Secc. 4.5) Demuéstrese que la ecuación (13) también es válida cuando μ es impar, tal que los valores posibles de A_k son $0, \pm 2A, \pm 4A, \ldots, \pm(\mu - 1)A$.

4.33 (Secc. 4.5) Un sistema digital polar tiene $\rho = 220$. ¿Cuál es el valor mayor permitido de μ tal que $P_e \leq 10^{-5}$?

4.34 (Secc. 4.5) Un sistema binario polar con 20 repetidores no regenerativos tiene $P_e = 10^{-4}$.

(*a*) Encuéntrese $(S/N)_1$.

(*b*) ¿En qué factor se debe incrementar r sin incrementar P_e si los repetidores se hacen regenerativos y el ancho de banda se modifica de manera apropiada?

5

MODULACIÓN LINEAL

Los diferentes casos de modulación ya quedaron asentados (Cap. 1) con una descripción cualitativa del proceso. Resumiendo brevemente: la *modulación* es la alteración sistemática de una forma de onda, conocida como la *portadora*, de acuerdo a las características de otra forma de onda, la señal moduladora o mensaje. La meta fundamental es producir una onda modulada que conduzca información, cuyas propiedades sean las más adecuadas al trabajo de comunicación dado.

Se inicia ahora un examen cuantitativo y un análisis de los sistemas de modulación, el *cómo* y el *porqué*; Este tema constituye el meollo del estudio de la comunicación y será dividido en cuatro partes: la *modulación lineal* (Cap. 5), la *modulación exponencial* (Cap. 6), el *ruido en la modulación de CW* (Cap. 7) y la *modulación de pulsos* (Cap. 8). La división material, hecha por conveniencia, de ninguna manera impide la claridad, aunque se encontrarán algunas repeticiones. En particular, tanto la modulación lineal como la exponencial, son tipos de *modulación de onda continua* (CW); es decir, que la portadora es una sinusoide. En consecuencia, hay similitudes inevitables de técnicas e instrumentación. La modulación de pulsos, en la cual la portadora es un tren de pulsos periódico, se puede asemejar a la modulación lineal, a la modulación exponencial o a ninguna, dependiendo del tipo. Por añadidura, se ha aplicado con gran éxito la moderna tecnología de pulsos a la generación y detección de la modulación de CW; y a la inversa, muchos sistemas de modulación de pulsos incluyen la modulación de CW para la etapa final en la transmisión.

FIGURA 5.1
Una señal pasabanda.

Como tema específico primario, diremos que la modulación lineal es, en esencia, la traslación directa de frecuencia del espectro del mensaje. La modulación de doble banda lateral (DSB) es eso precisamente. Las modificaciones menores del espectro trasladado dan lugar a la modulación de amplitud convencional (AM), a la modulación de banda lateral única (SSB) o a la modulación de banda lateral residual (VSB). Cada una de estas variantes tiene sus propias ventajas expresas y sus aplicaciones prácticas significativas. A cada una se le dará la debida importancia en este capítulo, incluyendo temas como formas de onda y espectros, métodos de detección, transmisores y receptores. El capítulo empieza con un examen general de las señales pasabanda y los sistemas pertinentes a todas las formas de modulación de CW.

5.1 SEÑALES Y SISTEMAS PASABANDA

Una señal pasabanda es la señal cuyo espectro se concentra en la vecindad de alguna frecuencia $f_c \neq 0$, conocida como la frecuencia portadora. Por lo tanto, como se muestra en la figura 5.1, la forma de onda se ve como una sinusoide de frecuencia f_c con cambios lentos de envolvente y de fase. Se escribe formalmente

$$v(t) = R(t)\cos[\omega_c t + \phi(t)] \qquad R(t) \geq 0 \tag{1}$$

donde $R(t)$ es la *envolvente* y $\phi(t)$ la *fase*, siendo ambas, funciones del tiempo. Por definición, la envolvente es *no negativa*; es decir, las "amplitudes" negativas se absorben en la fase al agregar $\pm$ 180°; así pues, la línea punteada de la figura 5.1 es $R(t)$.

La figura 5.2*a* es una representación fasorial de $v(t)$ como un vector en el plano complejo cuya longitud es igual a $R(t)$ y cuyo ángulo es igual a $\omega_c t + \phi(t)$. Pero el término angular $\omega_c t$ representa una rotación permanente en sentido contrario a las manecillas del reloj a f_c revoluciones por segundo y casi da lo mismo que se suprima, obteniéndose la figura 5.2*b*. Esta representación fasorial, que se empleará en forma regular en lo sucesivo, relaciona a la figura 5.2*a* de la manera siguiente: Si se fija el origen de la figura 5.2*b* y se gira la figura completa en sentido contrario a las manecillas del reloj a la rapidez f_c, se obtiene la figura 5.2*a*.

Una inspección más a fondo de la figura 5.2*b* sugiere otra forma de escribir a $v(t)$. Si se define

$$v_i(t) \triangleq R(t)\cos\phi(t) \qquad v_q(t) \triangleq R(t)\,\mathrm{sen}\,\phi(t) \tag{2}$$

FIGURA 5.2

entonces

$$v(t) = v_i(t)\cos\omega_c t - v_q(t)\,\mathrm{sen}\,\omega_c t \tag{3}$$

conocida como la descripción de la *portadora en cuadratura* de una señal pasabanda para distinguirla de la *descripción de envolvente y de fase*, ecuación (1); $v_i(t)$ y $v_q(t)$ se conocen como las *componentes en fase y en cuadratura*, respectivamente. La designación de portadora en cuadratura procede del hecho de que los dos términos de la ecuación (3) se pueden representar por medio de fasores, el segundo de los cuales está a un ángulo de + 90° en relación con el primero (véase Fig. 5.2*b*). El signo menos en la ecuación (3) puede ocasionar confusión, pero el lector puede justificarlo a partir de la ecuación (1) al desarrollar cos $[\omega_c t + \phi(t)]$

Ambas descripciones de señal pasabanda son útiles, pero la última presenta ventajas para la interpretación del dominio de la frecuencia. De manera específica, si $V_i(f) = \mathscr{F}[v_i(t)]$, etc., entonces, aplicando el teorema de la traslación de frecuencia, ecuación (22) (Sec. 2.3), a la ecuación (3) se obtiene

$$V(f) = \frac{1}{2}[V_i(f-f_c) + V_i(f+f_c)] + \frac{j}{2}[V_q(f-f_c) - V_q(f+f_c)] \tag{4}$$

puesto que $-$ sen $\omega_c t = \cos(\omega_c t + 90°)$ y $e^{\pm j90°} = \pm j$. Se tiene una implicación inmediata de la ecuación (4), ello si $v(t)$ es en efecto una señal pasabanda, entonces $v_i(t)$ y $v_q(t)$ son señales *pasabajas* y de banda limitada en $W < f_c$ (¿por qué?). La descripción de envolvente y de fase no se convierte de manera expedita al dominio de la frecuencia dado que, a partir de la ecuación (2) o de la figura 5.2*b*

$$R(t) = \sqrt{v_i^2(t) + v_q^2(t)} \qquad \phi(t) = \arctan\frac{v_q(t)}{v_i(t)} \tag{5}$$

las cuales no son transformables a expresiones de Fourier.

Señales y sistemas equivalentes pasabajas ★

Considérese la función de frecuencia pasabanda $V(f)$ (Fig. 5.3*a*) Su *equivalente pasabajas* se define

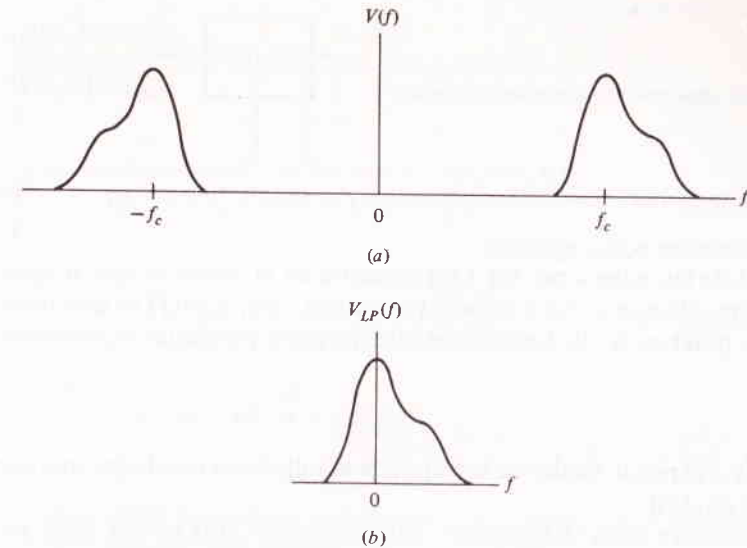



FIGURA 5.3
(*a*) Función pasabanda de la frecuencia $V(f)$; (*b*) función de la frecuencia equivalente pasabajas $V_{LP}(f)$.

$$V_{LP}(f) \triangleq V(f+f_c)u(f+f_c) = \begin{cases} V(f+f_c) & f > -f_c \\ 0 & f < -f_c \end{cases} \tag{6a}$$

la cual no es otra cosa que la porción de frecuencia positiva de $V(f)$ trasladada hacia abajo f_c unidades, (Fig. 5.3*b*.) La ecuación (6*a*) es una transformación de pasabanda a pasabajas y resulta, por lo general, una función no hermitiana. Para subrayar esta observación, la función del tiempo equivalente pasabajas es

$$v_{LP}(t) = \frac{1}{2}R(t)e^{j\phi(t)} = \frac{1}{2}v_i(t) + \frac{j}{2}v_q(t) \tag{6b}$$

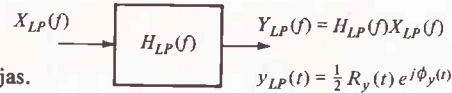
la cual es una función del tiempo *compleja*. Para deducir esta expresión, se empieza con la ecuación (4) y se sustituye f por $f + f_c$, dando $V(f + f_c) = (½)[V_i(f) + V_i(f + 2f_c] + (j/2)[V_q(f) - V_q(f + f2_c)]$. Entonces, puesto que $V_i(f)$ y $V_q(f)$ son funciones pasabajas, $V(f + f_c)u(f + f_c) = (½)V_i(f) + (j/2)V_q(f)$ y la transformación inversa conduce a la ecuación (6*b*).

Dado $v_{LP}(t)$, la función del tiempo pasabanda $v(t)$ se escribe con facilidad, ya sea en la forma envolvente y fase o en la de portadora en cuadratura, empleando

$$v(t) = 2\,\mathrm{Re}\,[v_{LP}(t)e^{j\omega_c t}] \tag{7a}$$

lo cual procede de la comparación de la ecuación (6*b*) con las ecuaciones (1) y (3). De manera similar, la función de la frecuencia pasabanda se relaciona con su equivalente pasabajas por medio de

FIGURA 5.4
Análisis de pasabanda empleando equivalentes pasabajas.

$X_{LP}(f)$ → $H_{LP}(f)$ → $Y_{LP}(f) = H_{LP}(f)X_{LP}(f)$, $y_{LP}(t) = \frac{1}{2}R_y(t)e^{j\phi_y(t)}$

$$V(f) = V_{\text{LP}}(f - f_c) + V^*_{\text{LP}}(-f - f_c) \tag{7b}$$

dejándose su deducción como ejercicio.

La utilidad de las ecuaciones (6) y (7) descansa en el hecho de que se aplican tanto a *sistemas* pasabanda como a señales pasabanda. Así, si $H(f)$ es una función de transferencia pasabanda, la función de transferencia pasabajas equivalente es simplemente

$$H_{\text{LP}}(f) = H(f + f_c)u(f + f_c) \tag{8}$$

De manera similar, se puede hablar de la respuesta al impulso a pasabajas equivalente $h_{LP}(t) = \mathscr{F}^{-1}[H_{LP}(f)]$.

Empleando ahora estas definiciones, supóngase que $x(t)$ es una señal pasabanda a la entrada de un sistema pasabanda. Si $X(f) = \mathscr{F}[x(t)]$, entonces el espectro de salida es $y(f) = H(f)X(f)$, como es usual. De manera alternativa, se puede demostrar que†

$$Y_{\text{LP}}(f) = H_{\text{LP}}(f)X_{\text{LP}}(f) \tag{9a}$$

$$y_{\text{LP}}(t) = h_{\text{LP}} * x_{\text{LP}}(t) \tag{9b}$$

y así se puede resolver el problema del análisis de pasabanda en términos de los equivalentes de pasabajas, lo cual es a menudo un trabajo más fácil. Así, por ejemplo, la envolvente y la fase de la señal de salida pasabanda $y(t)$ son

$$R_y(t) = 2|y_{\text{LP}}(t)| \qquad \phi_y(t) = \arg[y_{\text{LP}}(t)] \tag{10}$$

La figura 5.4 resume este método.

EJERCICIO 5.1 Dedúzcase la ecuación (7*b*) a partir de la (7*a*). (*Sugerencia*: Para cualquier cantidad compleja Z, $2\text{Re}[Z] = Z + Z^*$.)

Ejemplo 5.1 Retardo de portadora y de envolvente ★

Sea $x(t) = R_x(t)\cos\omega_c t$ una señal pasabanda aplicada a un sistema pasabanda que tiene relación de amplitudes constante $|H(f)| = K$ pero de fase *no lineal* $[H(f)] = \theta(f)$ sobre el intervalo de frecuencia de $x(t)$. Puesto que $H(f) = Ke^{j\theta}(f)$,

$$H_{\text{LP}}(f) = Ke^{j\theta(f + f_c)}u(f + f_c) \tag{11}$$

Si las no linealidades de fase no son muy severas, se puede desarrollar $\theta(f + f_c)$ en serie de Taylor, considerando sólo los dos primeros términos, así:

† Véase Frederick y Carlson (1971, Cap. 7).

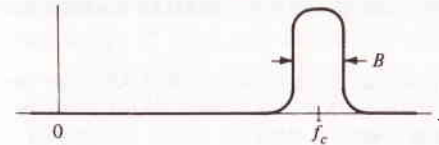



FIGURA 5.5
Respuesta de frecuencia típica de pasabanda.

$$\theta(f + f_c) = \theta(f_c) + f\left.\frac{d\theta(f)}{df}\right|_{f=f_c} + \cdots \approx -\omega_c t_c - \omega t_R \tag{12}$$

en la cual se han introducido

$$t_c \triangleq -\frac{\theta(f_c)}{2\pi f_c} \qquad t_R \triangleq -\frac{1}{2\pi}\left.\frac{d\theta(f)}{df}\right|_{f=f_c} \tag{13}$$

Estos parámetros se conocen como *retardo de la portadora* y *retardo de la envolvente*, respectivamente, por razones que pronto describiremos.

Insertando las ecuaciones (12) y (11) en la (9), se tiene

$$Y_{\text{LP}}(f) = Ke^{-j\omega_c t_c}X_{\text{LP}}(f)e^{-j\omega t_R}u(f + f_c)$$

así

$$y_{\text{LP}}(t) = Ke^{-j\omega_c t_c}x_{\text{LP}}(t - t_R) = Ke^{-j\omega_c t_c}(\tfrac{1}{2})R_x(t - t_R)$$

donde se ha empleado el teorema del retardo en el tiempo y anotado que $x_{\text{LP}}(t) = (\tfrac{1}{2})R_x(t)$ puesto que, por suposición, $\phi x(t) = 0$. Por último, de las ecuaciones (10) y (7*a*),

$$\begin{aligned} y(t) &= 2\,\text{Re}\,[Ke^{-j\omega_c t_c}(\tfrac{1}{2})R_x(t - t_R)e^{j\omega_c t}] \\ &= KR_x(t - t_R)\cos\omega_c(t - t_c) \end{aligned} \tag{14}$$

Por lo tanto, como implica lo anterior, la envolvente de salida se ha retardado en t_R mientras que el retardo de la portadora es t_c. Y puesto que t_R es independiente de la frecuencia, por lo menos para hacer mayor la aproximación involucrada, la envolvente no ha experimentado distorsión por retardo.

En ciertas aplicaciones —conocidas como sistemas de multicanalización por división en frecuencia— se transmiten varias señales pasabanda en diferentes portadoras sobre un canal único. Se emplean gráficas de $d\theta/df$ contra f como una ayuda en la evaluación de las características de retardo de la envolvente del canal. Si esta curva no es razonablemente plana sobre un intervalo de frecuencia propuesto, resultará una excesiva distorsión de envolvente.

Frecuencias de portadora y anchos de banda

Un sistema pasabanda analógico tiene la misma forma general de un sistema de banda base (Fig. 4.14) excepto que el transmisor, el canal y el receptor son unidades pasabanda. Así, en la figura 5.5, el ancho de banda B queda en forma nominal

Tabla 5.1 FRECUENCIAS PORTADORAS SELECCIONADAS Y ANCHO DE BANDA NOMINAL

	Banda de frecuencia	Frecuencia portadora	Ancho de banda
LW	Radio de onda larga	100 kHz	2 kHz
SW	Radio de onda corta	5 MHz	100 kHz
	VHF	100 MHz	2 MHz
	Microondas	5 GHz	100 MHz
	Ondas milimétricas	100 GHz	2 GHz
	Ópticas	5×10^{14} Hz	10^{13} Hz

centrado en la frecuencia portadora f_c. Aunque los parámetros B y f_c parecerán no estar relacionados, los factores prácticos hacen que lo estén, con mayor o menor desenvoltura.

Por ejemplo, las antenas en un sistema de radio producen una distorsión considerable a menos que el intervalo de frecuencia sea pequeño en comparación con la frecuencia portadora. Y en forma más general, el diseñar un amplificador pasabanda con una respuesta de frecuencia sin distorsión, se hace casi imposible si B es muy grande o muy pequeño comparado con f_c. Como una regla burda, el *ancho de banda fraccionario* B/f_c debe estar limitado por

$$0.01 < \frac{B}{f_c} < 0.1 \tag{15}$$

de otra manera habrá distorsión de la señal más allá del alcance de los igualadores prácticos. En forma incidental, conforme se aplica al factor de calidad Q de circuitos resonantes sencillos, la expresión (15) resulta $10 < Q < 100$ puesto que $B = f_r/Q$ (Fig. 2.12*a*).

Como una consecuencia inmediata de la ecuación (15) se puede decir que *anchos de banda grandes requieren frecuencias portadoras altas*. Este efecto ya se analizó brevemente (Cap. 1) y se recalca en la tabla 5.1, en la cual se listan las frecuencias portadoras seleccionadas y sus anchos de banda nominales $B \approx 0.02 f_c$. Para estar seguros, se pueden lograr anchos de banda más grandes, pero a costos sustancialmente mayores. Como una consecuencia más allá de la ecuación (15), los términos *pasabanda* y *banda angosta*, en forma virtual son sinónimos.

Transmisión de pulsos pasabanda

Por último, considérese el caso de una señal de pulsos pasabanda o modulada, de la forma

$$v(t) = p(t) \cos(\omega_c t + \theta) \tag{16}$$

donde $p(t)$ es una forma de pulso de banda base y θ es una constante. Se aprecia en forma inmediata, a partir del teorema de la modulación que el espectro de $v(t)$ está centrado en $f = f_c$ y tiene *dos veces* el ancho espectral de $p(t)$, como se ilustra en el ejemplo 2.7, (Sec. 2.3). Se sabe también que si $p(t)$ tiene una duración τ, entonces un análisis de pulso exitoso en banda base requiere un sistema pasabajas de ancho de banda $B_{LP} \geq \frac{1}{2}\tau$, por la ecuación (3), (Sec. 4.4).

Situándolos juntos dos a dos, se desprende que una solución acertada de $v(t)$ requiere un sistema pasabanda con ancho de banda $B = 2B_{LP} \geq 1/\tau$. En consecuencia, nuestra regla sencilla para la relación entre la duración de salida y el ancho de banda del sistema de pulsos pasabanda es

$$\tau_{\text{mín}} \geq \frac{1}{B} \tag{17a}$$

y, por medio de un razonamiento similar para el tiempo de subida de salida,

$$\tau_{r\,\text{mín}} \geq \frac{1}{B} \tag{17b}$$

Estas ecuaciones han servido durante mucho tiempo de guías burdas, pero útiles, en el trabajo de radar y campos relativos.

5.2 MODULACIÓN DE DOBLE BANDA LATERAL: AM Y DSB

Históricamente, el primer tipo de modulación en desarrollarse fue la modulación de amplitud convencional, la familiar AM de radiodifusión normal. Se reservará el término *modulación de amplitud* para este tipo específico, y el de *modulación lineal* para la clase general. La DSB (doble banda lateral) difiere de la AM en la *supresión de la portadora*, una alteración menor que tiene importantes repercusiones. Pero antes de entrar en detalles, se deben asentar unas cuantas convenciones y detalles de nomenclatura.

Convenciones de mensaje

El análisis de la transmisión en banda base analógica tuvo como base un *mensaje arbitrario* o conjunto de mensajes designado por $x(t)$. Siempre que sea posible, se seguirá adoptando esta práctica, aquí y en capítulos subsecuentes. También se seguirá suponiendo que $x(t)$ es aproximadamente de banda limitada en W, conocido como el *ancho de banda del mensaje*. Así,

$$\left.\begin{aligned} G_x(f) &\approx 0 \\ X(f) &\approx 0 \end{aligned}\right\} \quad |f| > W \tag{1}$$

La transformada de Fourier $X(f) = \mathscr{F}[x(t)]$ será de utilidad en muchos casos si el mensaje es una señal de energía, y la figura 5.6*a* será un espectro de mensaje representativo. La figura 5.6*b* muestra una densidad espectral de energía típica para una señal de voz e indica que W se debe considerar en la vecindad de los 3 a los 4 kHz.

Yendo más lejos, por conveniencia matemática, los mensajes se deben escalar o normalizar para tener una magnitud que no exceda de la unidad, es decir,

$$|x(t)| \leq 1 \tag{2}$$

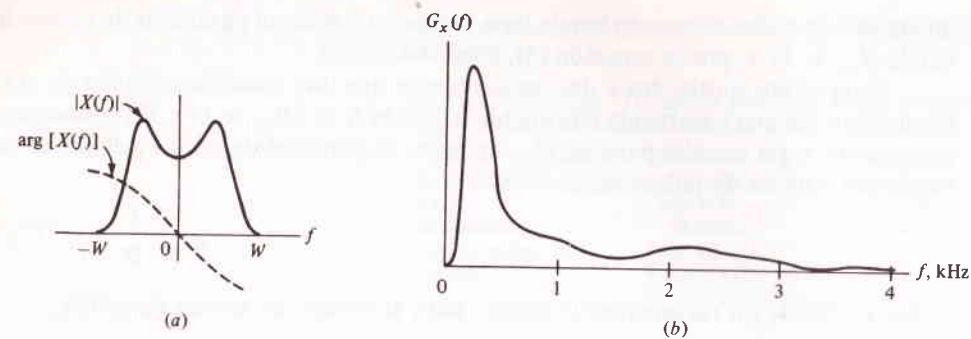



FIGURA 5.6
(a) Espectro de mensaje representativo; (b) densidad espectral típica de energía para señales de voz.

En consecuencia, cuando se considera la potencia promedio,

$$\langle x^2(t)\rangle = \overline{x^2} \leq 1 \tag{3}$$

suponiendo que la fuente es ergódica. Nótese, que nos moveremos libremente entre los modelos de señales de energía determinísticos y los de señales de potencia aleatorios que sean más apropiados al problema que nos ocupe.

En forma ocasional, un análisis con alguna $x(t)$ arbitraria, vendrá a ser difícil si no que imposible. Para tales situaciones es necesario recurrir a alguna señal moduladora específica, por lo general la simple sinusoide,

$$x(t) = A_m \cos 2\pi f_m t \qquad A_m \leq 1 \qquad f_m < W \tag{4}$$

a la cual se le conoce como *modulación de tono*. En ésta, será suficiente un espectro de líneas de frecuencia positiva para su estudio en el dominio de la frecuencia. Aunque la modulación de tono puede parecer indebidamente simplicada, la aproximación tiene ventajas definidas. Por un detalle, los tonos son a menudo las únicas señales fáciles de manejar en problemas complejos; facilitan el cálculo de espectros, potencias promedio, etc., que de otra manera serían prohibitivos. Más aún, si se puede encontrar la respuesta del sistema de modulación a una frecuencia específica en la banda del mensaje, se puede inferir la respuesta para todas las frecuencias en W.

Modulación en amplitud (AM)

El carácter distintivo de la AM radica en que la *envolvente* de la portadora modulada tiene la misma *forma* que la de la onda del mensaje. Esto se logra agregando el mensaje trasladado, en la proporción apropiada, a la portadora no modulada. En forma específica, la señal modulada es

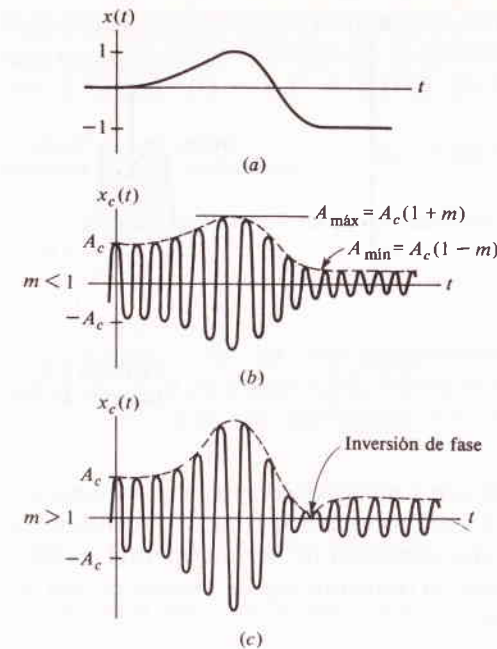



FIGURA 5.7
Formas de onda de AM. (a) $x(t)$; (b) $x_c(t)$ con $m < 1$; (c) $x_c(t)$ con $m > 1$.

$$\begin{aligned} x_c(t) &\triangleq A_c \cos \omega_c t + m x(t) A_c \cos \omega_c t \\ &= A_c[1 + m x(t)] \cos \omega_c t \end{aligned} \tag{5}$$

donde $A_c \cos \omega_0 t$ es la portadora no modulada, $f_c = \omega_c / 2\pi$ es la frecuencia portadora, y la constante m se conoce como *índice de modulación*. Puesto que A_c es la amplitud de la portadora no modulada, se puede pensar que la amplitud modulada sea una función lineal del mensaje, o sea,

$$A_c(t) = A_c[1 + m x(t)]$$

lo que subraya el significado de modulación *en amplitud*.

La figura 5.7 muestra una porción de un mensaje típico y la onda resultante de AM para dos valores de m. La envolvente tiene la forma de $x(t)$ siempre y cuando la frecuencia portadora sea mucho mayor que la rapidez de variación de $x(t)$, de otra manera no se puede visualizar su envolvente y no hay *inversiones de fase* en la onda modulada, es decir, la amplitud $A_c[1 + m x(t)]$ no se hace negativa. Así, se conserva la relación deseada entre la envolvente y el mensaje

$$f_c \gg W \qquad \text{y} \qquad m \leq 1 \tag{6}$$

Lo último procede de que $|x(t)| \leq 1$ por convención.

La condición de frecuencia portadora está, por supuesto, de acuerdo con la traslación de frecuencia, un aspecto de la modulación, aunque algunas veces resulta perjudicial. La condición $m \leq 1$ fija un límite superior de intensidad a que se puede mo-

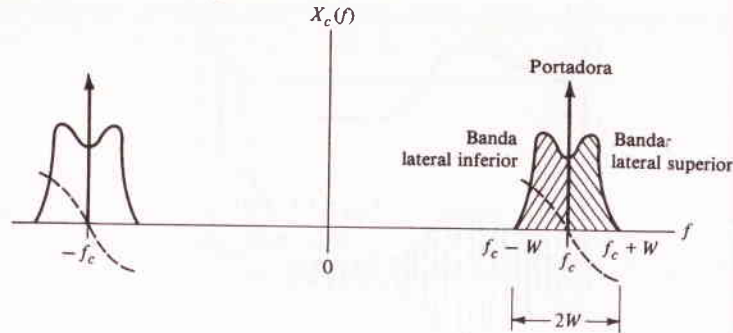



FIGURA 5.8
Espectro de AM.

dular la portadora. Con $m = 1$, conocido como modulación al ciento por ciento, la amplitud modulada varía entre 0 y $2A_c$. La *sobremodulación*, $m > 1$, ocasiona inversiones de fase en la portadora y con ello *distorsión de envolvente* (Fig. 5.7*c*).

Volviendo al dominio de la frecuencia, la transformada de Fourier de una onda de AM se encuentra con facilidad y es

$$X_c(f) = \frac{A_c}{2}[\delta(f-f_c) + \delta(f+f_c)] + \frac{mA_c}{2}[X(f-f_c) + X(f+f_c)] \qquad (7)$$

Si la ecuación (7) puede parecer imponente, el espectro está formado nada menos que del espectro del mensaje trasladado, y un par de impulsos en $\pm f_c$ representando a la propia portadora, (Fig. 5.8). Se deben notar con sumo cuidado dos propiedades del espectro de la AM.

1 Existe *simetría* en relación con la frecuencia portadora, siendo la amplitud par y la fase impar. La porción del espectro por encima de f_c se conoce como *banda lateral superior*, y la que está por debajo de f_c es la *banda lateral inferior*; de ahí la designación de modulación de amplitud de *doble banda lateral*.
2 El *ancho de banda de transmisión* b_T requerido para una señal de AM es exactamente *dos veces* el ancho de banda del mensaje, o sea

$$B_T = 2W \qquad (8)$$

Este resultado señala que la AM no es conveniente cuando se debe conservar el ancho de banda. Después de todo, el mensaje debería ser enviado en banda base con la mitad del ancho de banda de la AM.

Se encontrará que el ancho de banda de transmisión es uno de los parámetros característicos de los tipos de modulación; otro parámetro importante es la *potencia transmitida promedio* en la onda modulada, que se define como

$$S_T = \mathrm{E}[x_c^2(t)]$$

Para evaluar este promedio de ensamble, se agrega un ángulo aleatorio Θ a la portadora que $x_c(t)$ tenga la forma de un proceso aleatorio modulado, ecuación (11*a*), (Sec. 3.5), con $v(t) = A_c[1 + mx(t)]$. De donde se tiene

$$S_T = \tfrac{1}{2}\mathrm{E}[v^2(t)] = (1 + 2m\overline{x} + m^2\overline{x^2})\frac{A_c^2}{2}$$

Por lo que, suponiendo que el mensaje no tiene componente de CD,†

$$S_T = (1 + m^2\overline{x^2})\frac{A_c^2}{2} \qquad (9)$$

La ecuación (9) tiene una interpretación interesante en relación con el espectro de AM. A partir de la figura 5.8, S_T incluye la potencia de la componente de frecuencia de la portadora más dos bandas laterales simétricas. En consecuencia

$$S_T = \frac{A_c^2}{2} + m^2\overline{x^2}\frac{A_c^2}{2} = P_c + 2P_{SB}$$

donde la *potencia de la portadora* es

$$P_c = \tfrac{1}{2}A_c^2$$

y la *potencia por cada banda lateral*

$$P_{SB} = \frac{m^2\overline{x^2}P_c}{2} \le \tfrac{1}{2}P_c$$

El límite superior en P_{SB} procede de $m^2\overline{x^2} \le 1$. Así

$$P_c = S_T - 2P_{SB} \ge \tfrac{1}{2}S_T \qquad (10)$$

lo cual dice que por lo menos el 50% de la potencia transmitida total reside en la portadora. Pero haciendo referencia a la ecuación (5), se ve que el término portadora sólo es independiente del mensaje y no contiene ninguna información. Por lo que se concluye que una porción sustancial de la potencia S_T transmitida se "desperdicia" en la propia portadora. Se demostrará más adelante que la verdadera modulación de envolvente, y la sencillez de la *detección de envolvente*, dependen de esta potencia desperdiciada.

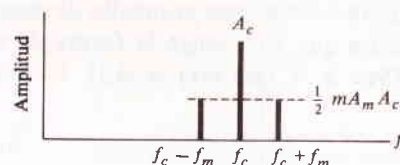
AM con modulación de tono

Haciendo $x(t) = Am \cos 2\pi f_m t$ se tiene la forma de onda de AM de tono modulado

$$x_c(t) = A_c(1 + mA_m \cos \omega_m t)\cos \omega_c t$$

$$= A_c \cos \omega_c t + \frac{mA_m A_c}{2}[\cos(\omega_c - \omega_m)t + \cos(\omega_c + \omega_m)t] \qquad (11)$$

† Por las razones analizadas (Sec. 5.5), los mensajes que tiene $\overline{x} \neq 0$ es raro que se transmitan en AM.

FIGURA 5.9
Espectro de líneas de AM, modulación de tono.

donde se emplea el desarrollo trigonométrico para el producto de dos cosenos. En la figura 5.9 se muestra el correspondiente espectro de líneas de frecuencia positiva.

Con la modulación de tono, una onda de AM se puede tratar como una *suma de fasores*, uno por cada línea espectral. En la figura 5.10 se muestra una manera especialmente informativa para la construcción del diagrama, donde los fasores correspondientes a cada banda lateral se agregan al extremo del fasor portadora. Puesto que el fasor portadora gira a f_c Hz, los fasores de las bandas laterales giran a velocidades de $\pm f_m$ *en relación* con la portadora. Nótese que mientras que las líneas de las bandas laterales son iguales y de fase correcta, la resultante de los fasores de las bandas laterales son *colineales* con el fasor de la portadora. Además, la magnitud de la resultante es la envolvente $R(t)$.

Esta observación conduce a una forma simple de estudiar de una manera cualitativa los efectos de las imperfecciones de transmisión, de la interferencia, etc. Por ejemplo, si se atenúa en forma muy severa la línea correspondiente a la banda lateral inferior (Fig. 5.11), la envolvente resultante es

$$R(t) = \left[\left(A_c + \frac{mA_mA_c}{2}\cos\omega_m t\right)^2 + \left(\frac{mA_mA_c}{2}\operatorname{sen}\omega_m t\right)^2\right]^{1/2}$$

$$= A_c\left[1 + \left(\frac{mA_m}{2}\right)^2 + mA_m\cos\omega_m t\right]^{1/2} \qquad (12)$$

de la cual se puede determinar la *distorsión de envolvente.*

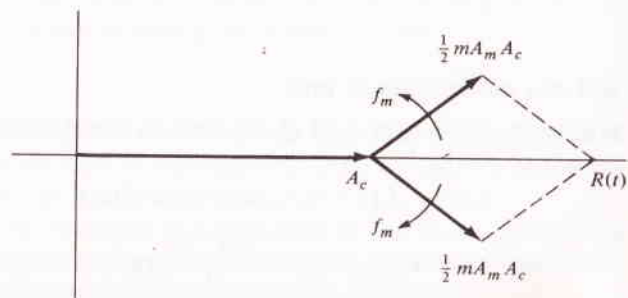



FIGURA 5.10
Diagrama fasorial para AM con modulación de tono.

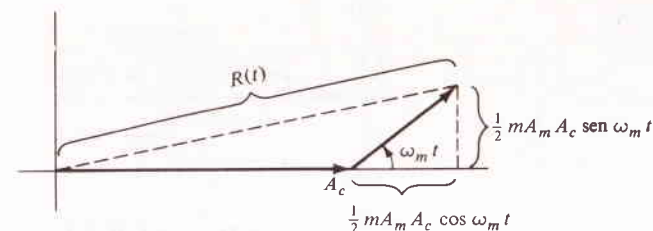



FIGURA 5.11

Modulación de doble banda lateral con portadora suprimida (DSB)

Como la componente portadora en AM es independiente del mensaje y viene a ser una potencia desperdiciada, se puede eliminar de la onda modulada. Esto ocasiona una modulación de amplitud de doble banda lateral con portadora suprimida, *modulación de doble banda lateral*, o DSB† por brevedad.

Por la supresión del término portadora y el ahora carente de sentido índice de modulación de la ecuación (5) se origina

$$x_c(t) \triangleq x(t)A_c\cos\omega_c t \qquad (13)$$

por la definición de DSB. En esto nada parecido a la AM, la onda modulada es cero en ausencia de modulación: $x_c(t) = 0$ cuando $x(t) = 0$. Puesto que se ha suprimido la portadora, la potencia transmitida promedio es

$$S_T = 2P_{SB} = \tfrac{1}{2}\overline{x^2}A_c^2 \qquad (14)$$

El espectro de la modulación de doble banda lateral (DSB) es simplemente el espectro de mensaje trasladado

$$X_c(f) = \frac{A_c}{2}[X(f-f_c) + X(f+f_c)] \qquad (15)$$

Por supuesto, el ancho de banda de transmisión no cambia en relación con el caso de la AM, o sea

$$B_T = 2W$$

La comparación de las ecuaciones (15) y (7) demuestra que la AM y la DSB son muy similares en el dominio de la frecuencia, pero la gráfica del dominio del tiempo es otra cosa. Como se ilustra en la figura 5.12, la envolvente de DSB no tiene la misma forma que la del mensaje, puesto que los valores negativos de $x(t)$ se reflejan en la *fase* de la portadora. Así, dondequiera que $x(t)$ cruza el nivel cero, la onda modulada tiene una *inversión de fase*. La recuperación completa del mensaje implica una

† También se emplean las abreviaturas DSB-SC y DSSC (en inglés).

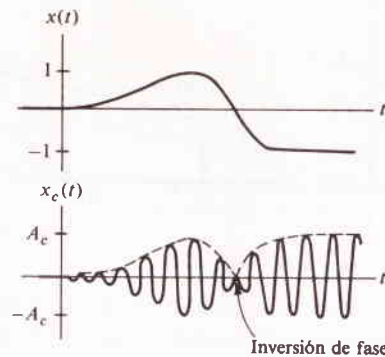



FIGURA 5.12
Formas de onda de modulación de doble banda lateral.

observación precisa de estas inversiones de fase, y un detector de envolvente no sería suficiente. Establecido de otra manera, la modualción de doble banda lateral, implica sólo modulación de "amplitud".

Esta consideración sugiere, y el análisis posterior lo confirma, que existe un compromiso entre la eficiencia de potencia y los métodos de demodulación. La modulación de doble banda lateral conserva potencia pero requiere de circuitos de detección complejos; y a la inversa, la demodulación de la AM es una simple detección de envolvente pero al precio de una potencia de transmisión mayor.

EJERCICIO 5.2 Demuéstrese que las representaciones de envolvente y fase y de portadora en cuadratura de AM (con $m \leq 1$) y de DSB son como sigue.

$$R_{AM}(t) = A_c[1 + mx(t)] \qquad R_{DSB}(t) = A_c|x(t)| \tag{16a}$$

$$\phi_{AM}(t) = 0 \qquad \phi_{DSB}(t) = \begin{cases} 0 & x(t) > 0 \\ 180° & x(t) < 0 \end{cases} \tag{16b}$$

$$v_{i_{AM}}(t) = R_{AM}(t) \qquad v_{i_{DSB}}(t) = A_c x(t) \tag{16c}$$

$$v_{q_{AM}}(t) = 0 \qquad v_{q_{DSB}}(t) = 0 \tag{16d}$$

EJERCICIO 5.3 Sea $x(t)$ la onda binaria aleatoria (Fig. 3.9) con $A = 1$. Bosquéjense las formas de onda moduladas resultantes si la modulación es en AM con $m = 0.5$, en AM con $m = 1.0$ y en DSB. Empléense estos bosquejos para explicar por qué algunas veces se dice que la DSB es una AM con una modulación al 200%.

5.3 MODULADORES Y TRANSMISORES

Se ha visto que se han creado nuevas frecuencias en el proceso de la modulación. El dispositivo que genera una onda de AM, un *modulador*, debe, por lo tanto, ser o *variable en el tiempo* o *no lineal*, puesto que los sistemas invariantes en el tiempo y lineales no pueden producir nuevas componentes de frecuencia.

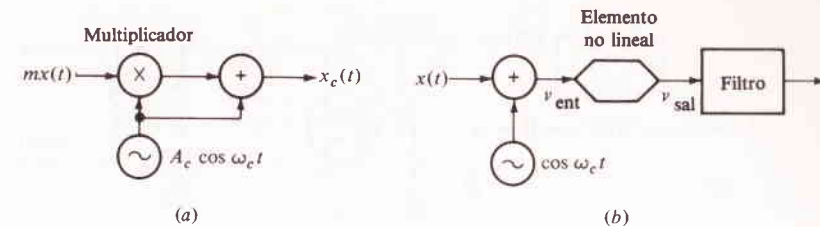



FIGURA 5.13
Moduladores de AM empleando: (*a*) multiplicador; (*b*) elemento no lineal.

La figura 5.13*a* ilustra un modulador posible, que se deriva en forma directa de la definición de la AM. Como se requiere de un multiplicador, el sistema se puede designar como *modulador de producto*. En realidad, recordando que de la multiplicación por $\omega_c t$ resulta una traslación de frecuencia, la operación producto resulta básica en los sistemas de modulación lineal. En la práctica, la multiplicación se hace con la ayuda de un elemento no lineal y filtraje, en un arreglo (Fig. 5.13*b*), donde la no linealidad puede ser de *ley de potencia* o *lineal y en forma fragmentaria* (conmutación).

Moduladores de ley de potencia

Tomando el caso de ley de potencia, considérese el dispositivo no lineal cuya característica de transferencia es

$$v_{salida} = a_1 v_{entrada} + a_2 v_{entrada}^2$$

Omitiendo los términos de orden superior, se le supone un dispositivo de ley *cuadrática*. Entonces, si $v_{in}(t) = x(t) + \cos \omega_c t$.

$$v_{entrada}(t) = x(t) + \cos \omega_c t,$$

$$v_{salida}(t) = a_1 x(t) + a_2 x^2(t) + a_2 \cos^2 \omega_c t + a_1 \left[1 + \frac{2a_2}{a_1} x(t)\right] \cos \omega_c t \tag{1}$$

El último término es la onda de AM deseada, con $A_c = a_1$ y $m = 2a_2/a_1$, siempre que pueda ser separada del resto.

La figura 5.14 muestra, como una posibilidad de separación, el espectro $V_{salida}(f) = \mathscr{F}[v_{salida}(t)]$ tomando a $X(f)$ como en la figura 5.6*a*. Nótese que el térmi-

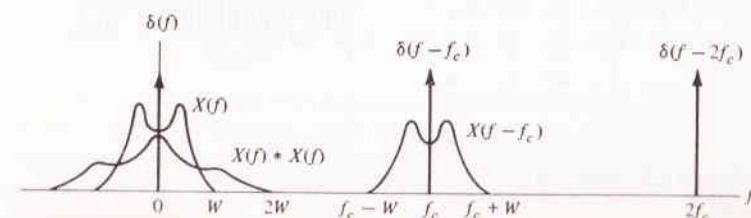



FIGURA 5.14

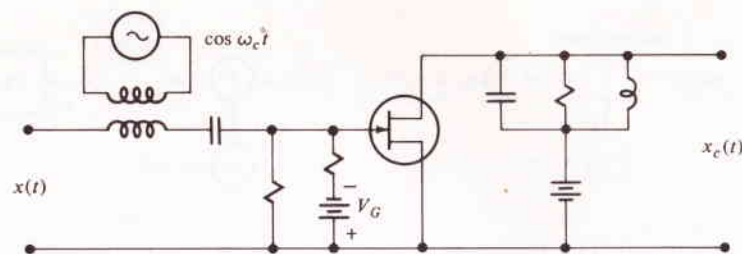



FIGURA 5.15
Circuito del modulador de AM que emplea un transistor de efecto de campo como un dispositivo de ley cuadrática.

no $x^2(t)$ de la ecuación (1) resulta ser $X * X(f)$, la cual es de banda limitada en $2W$. Por lo que, si $f_c > 3W$, no existe traslape espectral y con ello se puede lograr la separación requerida, por medio de un filtro pasabanda de ancho de banda $B_T = 2W$ centrado en f_c.

En la figura 5.15 se muestra el circuito completo de un modulador que emplea un transistor de efecto de campo (FET). La batería V_G sirve para polarizar al FET en su región de saturación donde tiene una característica de ley cuadrática, y el circuito RLC en paralelo es el filtro pasabanda.†

Como se necesita un filtraje de importancia en los moduladores de ley de potencia, se emplean principalmente para modulación de *bajo nivel*, es decir, a niveles de potencia menores que el valor transmitido. Se necesita entonces una amplificación lineal sustancial para elevar la potencia al valor S_T. Pero los amplificadores de potencia de RF de la linealidad que se necesitan presentan sus problemas y a menudo es mejor emplear la modulación de *alto nivel* si S_T va a ser grande.

Moduladores de conmutación

Los moduladores eficientes de alto nivel se arreglan o disponen de tal manera que nunca se desarrollen en forma completa productos de modulación indeseables que no necesiten filtraje. Esto se logra generalmente con la ayuda de un dispositivo *de*

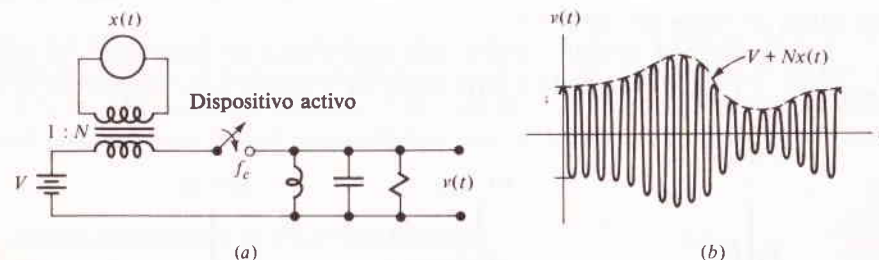



FIGURA 5.16
Amplificador clase C modulado por suministro de voltaje. (*a*) Circuito equivalente; (*b*) forma de onda de salida.

† En Clark y Hess (1971, Cap. 8) se da el análisis completo.

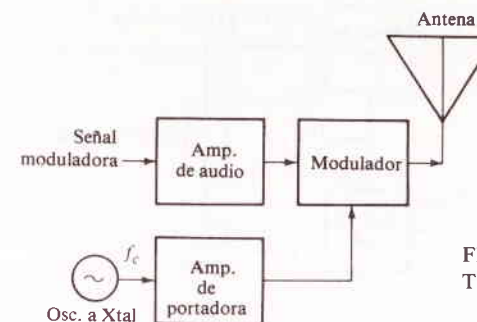



FIGURA 5.17
Transmisor de AM con modulación de alto nivel.

conmutación, cuyo análisis detallado se trata en el capítulo 8. Sin embargo, la operación básica del amplificador clase C modulado con suministro de voltaje, se comprende en forma rápida a partir de su circuito ideal equivalente y de las formas de onda, (Fig. 5.16).

El dispositivo activo, un tubo al vacío o un transistor, sirve como un conmutador impulsado a la frecuencia portadora, cerrando brevemente cada $1/f_c$. La carga, conocida como circuito *tanque*, se sintoniza para resonar a f_c, con lo que la acción de conmutación origina que el circuito tanque "oscile" en forma senoidal. El voltaje en la carga de estado permanente en ausencia de modulación es entonces $v(t) = V \cos \omega_c t$. Al agregar el mensaje a la fuente de voltaje, digamos por medio del transformador, se tiene $v(t) = [V + Nx(t)] \cos \omega_c t$, donde N es la relación de vueltas del transformador. Si V y N se proporcionan en forma correcta, se logra la modulación deseada sin una generación apreciable de componentes indeseables.

En la figura 5.17 se muestra el diagrama esquemático de un transmisor completo de AM para el caso de modulación de alto nivel. La onda portadora se genera por medio de un oscilador controlado a cristal, para asegurar la estabilidad de la frecuencia portadora. Como la modulación de alto nivel demanda señales de entrada potentes, tanto la portadora como el mensaje se amplifican antes de la modulación. La señal modulada se entrega luego directamente a la antena.

Moduladores balanceados

Los moduladores de DSB difieren de los de AM sólo en la supresión de la componente portadora; en teoría, esto necesita nada más y nada menos que un *multiplicador*. Como el lector puede verificar, el sistema de la figura 5.13*b* generará una DSB si el elemento no lineal es un dispositivo de ley cuadrática *perfecto*, tal que $v_{\text{salida}} = a v_{\text{entrada}}^2$.

De manera desafortunada, los dispositivos de ley cuadrática perfectos son raros, ya que en la práctica la DSB se obtiene empleando *dos* moduladores de AM dispuestos en una configuración balanceada para cancelar la portadora. La figura 5.18 muestra tal *modulador balanceado* en forma de diagrama a bloques. Suponiendo que los moduladores de AM son idénticos, salvo por el signo invertido de una entrada, las salidas son $A_c[1 + \frac{1}{2}x(t) \cos \omega_c t$ y $A_c[1 - \frac{1}{2}x(t)] \cos \omega_c t$. Restando una de

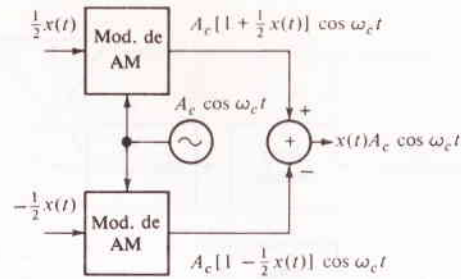



FIGURA 5.18
Modulador balanceado

la otra se obtiene $x_c(t) = x(t)A_c \cos \omega_c t$, conforme se necesita. En consecuencia, un modulador balanceado es un multiplicador. Se debe notar que si el mensaje tiene un *término* de CD, esa componente *no* se cancela en el modulador, por lo que aparecerá en la frecuencia portadora en la onda modulada.

En el capítulo 8 se examina un modulador balanceado que emplea la conmutación, bajo el encabezado de *modulación de troceador* ("chopper"). En la literatura respectiva se pueden encontrar otras realizaciones de circuitos.

EJERCICIO 5.4 Supóngase que los moduladores de AM (Fig. 5.18) se construyen con elementos no lineales idénticos que tienen $v_{\text{salida}} = a_1 v_{\text{entrada}} + a_2 v_{\text{entrada}}^2 + a_3 v_{\text{entrada}}^3$. Tomando una entrada arbitraria $x(t)$, demuéstrese que las señales de AM tienen distorsión por segunda armónica pero, no obstante ello, la salida final es DSB sin distorsión.

5.4 MODULACIÓN DE BANDA LATERAL SUPRIMIDA: SSB Y VSB (BANDA LATERAL ÚNICA Y BANDA LATERAL RESIDUAL)

La AM convencional implica un gran desperdicio, tanto de la potencia transmitida como del ancho de banda de transmisión. La supresión de la portadora subsana el primer defecto; la supresión de una banda lateral, por completo o en parte, reduce el segundo y conduce a la modulación de banda lateral única (SSB) o de banda lateral residual (VSB).

Modulación de banda lateral única (SSB)

Las bandas laterales superior e inferior de AM o de DSB se relacionan en forma exclusiva por la simetría en relación con la frecuencia portadora; si se da la amplitud y la fase de una, se puede reproducir la otra. En consecuencia, el ancho de banda de transmisión se puede recortar a la mitad con lo que se suprime una banda lateral lo mismo que la portadora. La eliminación total de la portadora y de una banda lateral del espectro de AM de la figura 5.8, produce la SSB, por lo que es obvio que

$$B_T = W \qquad S_T = P_{SB} = \tfrac{1}{4}\overline{x^2}A_c^2 \tag{1}$$

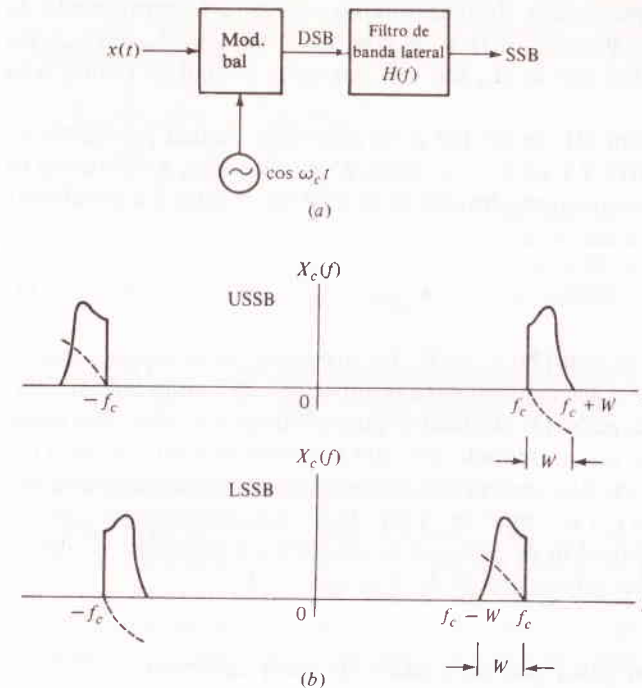



FIGURA 5.19
Modulación de banda lateral única. (*a*) modulador; (*b*) espectros.

La SSB se visualiza con rapidez en el dominio de la frecuencia: es la salida de un modulador balanceado procesado por medio de un *filtro de banda lateral* (Fig. 5.19*a*). El filtro es un circuito pasabanda que permite el paso de la banda lateral superior (USSB) o de la banda lateral inferior (LSSB), según sea el espectro resultante como se muestra en la figura 5.19*b*.

Por otra parte, la descripción en el dominio del tiempo de la forma de onda, es algo más difícil, salvo para el caso de la *modulación de tono*. Con la modulación de tono $x_c(t)$ es la sencillez misma; la supresión de la portadora y una línea de banda lateral en la figura 5.9 deja sólo la otra línea de banda lateral, es decir,

$$x_c(t) = \tfrac{1}{2}A_m A_c \cos(\omega_c \pm \omega_m)t \tag{2}$$

La frecuencia de la onda modulada estará, por lo tanto, a un lado de f_c en $\pm f_m$, en tanto que la amplitud es constante pero proporcional a la amplitud del tono.

Para una $x(t)$ arbitraria, en un análisis que se realiza adelante, se demuestra que

$$x_c(t) \triangleq \tfrac{1}{2}A_c[x(t)\cos\omega_c t \mp \hat{x}(t)\operatorname{sen}\omega_c t] \tag{3}$$

donde se toma el signo superior para la banda lateral superior y viceversa, y donde $\hat{x}(t)$ es la *transformada de Hilbert* de $x(t)$. En forma física, $\hat{x}(t)$ es simplemente $x(t)$

con sus componentes de frecuencia desplazados en $-90°$. Comprobando la ecuación (3) con modulación de tono, $\hat{x}(t) = A_m\cos(\omega_m t - 90°) = A_m$ sen $\omega_m t$, así $x_c(t) = \frac{1}{2}A_c(A_m \cos \omega_m t \cos \omega_c t \pm A_m$ sen $\omega_m t$ sen $\omega_c t)$, lo cual se reduce a la ecuación (2).

Al examinar la ecuación (3), se ve que es la expresión de una *portadora en cuadratura* con $v_i(t) = A_c x(t)/2$ y $v_q(t) = \pm A_c(t)/2$. Por lo tanto, a diferencia de la AM y la DSB, la componente en cuadratura de la SSB no es cero. La envolvente de la SSB es así,

$$R_{SSB}(t) = \frac{A_c}{2}\sqrt{x^2(t) + \hat{x}^2(t)} \tag{4}$$

en tanto que $\phi_{SSB}(t) = \pm \arctan [\hat{x}(t)/x(t)]$. La presencia de la componente en cuadratura de la SSB vuelve a ser esencial para la supresión de banda lateral. Para captar esto, nótese que la ecuación (3) también se puede interpretar como *dos ondas* DSB en cuadratura de fase, una modulada por $x(t)$ y la otra por $\hat{x}(t)$. Puesto que $x_c(t)$ es banda lateral *única*, las dos ondas DSB deben ser tales que sus bandas laterales se cancelan de un lado de f_c y se suman en el otro lado. Esta observación conduce a un método alterno de generación de SSB que se describirá brevemente. Primero, sin embargo, se aclararán las implicaciones de la ecuación (4).

Ejemplo 5.2 Banda lateral única con modulación de onda cuadrada

Siempre que la señal moduladora $x(t)$ en SSB tenga transiciones de escalón, el término en cuadratura $\hat{x}(t)$ presenta picos agudos; en consecuencia, por la ecuación (4) la envolvente tendrá también picos agudos, conocidos algunas veces como *cuernos*. La demostración de este efecto en una modalidad sencilla, se logra considerando que la señal moduladora esté dada en forma aproximada por

$$x(t) \approx \cos \omega_m t - \tfrac{1}{3} \cos 3\omega_m t + \tfrac{1}{5} \cos 5\omega_m t \tag{5a}$$

la cual es la señal de prueba empleada (Sec. 4.2, Fig. 4.4). Se podría escribir entonces $x_c(t)$ como la suma de tres términos, cada uno de los cuales forma la ecuación (2). Pero $\hat{x}(t)$ se encuentra con facilidad en este caso por el corrimiento de fase de $x(t)$, es decir,

$$\hat{x}(t) \approx \text{sen}\, \omega_m t - \tfrac{1}{3} \text{sen}\, 3\omega_m t + \tfrac{1}{5} \text{sen}\, 5\omega_m t \tag{5b}$$

lo cual se aproxima a una onda *triangular*, conforme a la gráfica de la figura 4.7.

Sustituyendo las ecuaciones (5*a*) y (5*b*) en la (4) y desarrollando en forma rutinaria se tiene el resultado

$$R(t) = \frac{A_c}{2}\left[\left(1 + \frac{1}{9} + \frac{1}{25}\right) - \frac{4}{5}\cos 2\omega_m t + \frac{2}{5}\cos 4\omega_m t\right]^{1/2} \tag{6}$$

expresión que se bosqueja en la figura 5.20 con $x(t)$. Los cuernos de la envolvente son visibles claramente y de haber tomado una onda cuadrada ideal, esos cuernos habrían sido picos infinitos. Es obvio que esto no presagia nada bueno para efectos de la transmisión de pulsos por medio de la SSB.

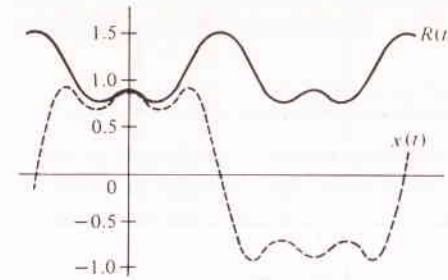



FIGURA 5.20
La envolvente de una forma de onda de SSB ($\cos A_c = 2$) cuando $x(t)$ se aproxima a una onda cuadrada.

EJERCICIO 5.5 Dibújese el espectro de líneas de un sólo lado y encuéntrense P_{SB} y S_T para AM ($m = 1$), DSB y USSB cuando $x(t) = \frac{1}{3} \cos \omega_m t + \frac{2}{3} \cos 4\omega_m t$.

Transformadas de Hilbert y señales analíticas

Un análisis completo de la SSB requiere del estudio de las transformadas de Hilbert y los conceptos relacionados con las señales analíticas. Éstas a su vez se relacionan con el *corrimiento de fase en cuadratura*, que se tomará como punto de partida.

Considérese una red o filtro cuya función no es otra que desplazar la fase de todas la componentes de frecuencia positiva en $-90°$ y todas las componentes de frecuencia negativa en $+90°$. Puesto que un corrimiento de fase de $\pm 90°$ equivale a multiplicar por $e^{\pm j90°} = \pm j$, la función de transferencia es como la que se dibuja en la figura 5.21 y se puede escribir como

$$H_Q(f) = -j\, \text{sgn}\, f \tag{7}$$

donde sgn f, que se lee "signum f", es la *función signo*

$$\text{sgn}\, \lambda \triangleq \begin{cases} 1 & \lambda > 0 \\ -1 & \lambda < 0 \end{cases} \tag{8}$$

Si $X(f)$ es el espectro de entrada a la red, el espectro de salida es $-j$ sgn $fX(f)$.

Pero el corrimiento de fase en cuadratura no se interpreta con facilidad en el dominio del tiempo, y para este propósito es necesario volver a la respuesta al impulso del desplazador de fase en cuadratura. Tomar la transformada inversa de la ecuación (8) requiere una integración de contorno o una operación de límites. Se ob-

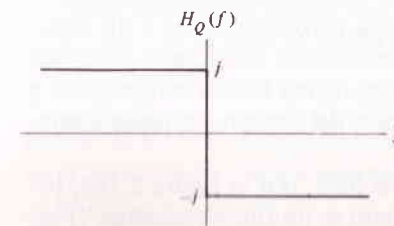



FIGURA 5.21
La función de transferencia de un desplazador de fase en cuadratura.

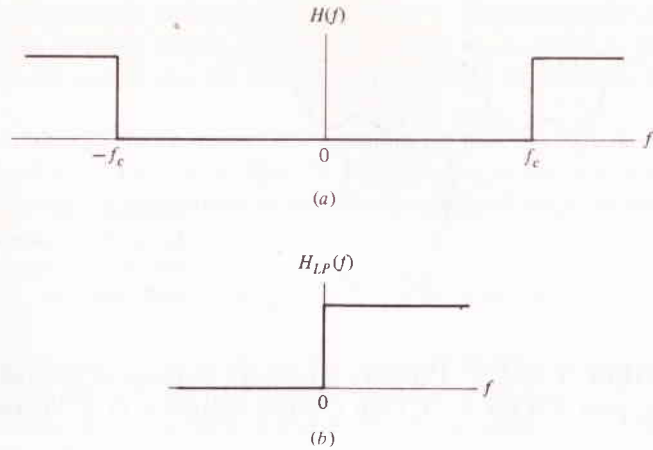



FIGURA 5.22
(a) Filtro de banda lateral USSB ideal; (b) filtro equivalente pasabajas.

tiene una aproximación más sencilla a la respuesta al impulso por medio de teoremas transformados y un poco de ingenio. Recuérdese el teorema de la diferenciación, $dv(t)dt \longleftrightarrow j2\pi fV(f)$. Su expresión dual es $-j2\pi tv(t) \longleftrightarrow dV(f)/df$, y así $v(t) = (j/2\pi t)\mathscr{F}^{-1}[dV(f)/df]$. Con la ayuda
de la figura 5.21, se puede diferenciar $H_a(f) = -j \operatorname{sgn} f$, obteniéndose $dH_Q(f)/df = -j2\delta(f)$. Por lo que la correspondiente respuesta al impulso es

$$h_Q(t) = \frac{j}{2\pi t}\mathscr{F}^{-1}[-j2\,\delta(f)] = \frac{1}{\pi t}$$

Generalizando y aplicando la dualidad, se deducen los pares transformados

$$\frac{1}{\pi t} \leftrightarrow -j \operatorname{sgn} f \qquad (9a)$$

$$\operatorname{sgn} t \leftrightarrow \frac{-j}{\pi f} \qquad (9b)$$

Se puede decir ahora que si $x(t)$ es la entrada a un desplazador de fase en cuadratura, la función del tiempo de salida es

$$\hat{x}(t) = [x(t)] * \left[\frac{1}{\pi t}\right] = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{x(\lambda)}{t-\lambda}\,d\lambda \qquad (10)$$

Se definió a $\hat{x}(t)$ como transformada de Hilbert de $x(t)$, aunque algunos autores emplean la forma negativa de la ecuación (10) como la correspondiente a un corrimiento de fase de $+90°$. A $\hat{x}(t)$ también se le conoce como la *conjugada armónica* de $x(t)$. Nótese que la transformación de Hilbert es en forma básica convolución y no da lugar a un cambio de *dominio*; si x es una función del tiempo, entonces $\hat{x}$ también lo es.

Aplicando estos resultados a la generación de la SSB, por la figura 5.19a, tómese el caso de la USSB donde el filtro de banda lateral es un filtro pasaaltas, (Fig.

5.22a). El filtro pasabajas equivalente (Fig. 5.22b), se encuentra a partir de la ecuación (8), (Sec. 5.1), como

$$H_{LP}(f) = u(f) = \frac{1}{2}(1 + \operatorname{sgn} f) = \frac{1}{2} + \frac{j}{2}(-j \operatorname{sgn} f) \qquad (11a)$$

donde la muy extraña tercera expresión anticipa el siguiente paso, o sea, encontrar la respuesta al impulso equivalente del pasabajas. De manera específica, como se desprende de las ecuaciones (11a) y (9a),

$$h_{LP}(t) = \mathscr{F}^{-1}[H_{LP}(f)] = \frac{1}{2}\delta(t) + \frac{j}{2}\frac{1}{\pi t} \qquad (11b)$$

Ahora, puesto que la entrada al filtro de banda lateral es una onda DSB, $A_c x(t) \cos \omega_c t$, la entrada equivalente del pasabajas es simplemente $(A_c/2)x(t)$ y la salida es

$$y_{LP}(t) = [h_{LP}(t)] * \left[\frac{A_c}{2}x(t)\right]$$

$$= \frac{A_c}{4}[x(t) + j\hat{x}(t)] \qquad (12)$$

puesto que $[\delta(t)]*[x(t)] = x(t)$ en tanto que $[1/\pi t] * [x(t)] = \hat{x}(t)$. Por último, la transformación de pasabajas a pasabanda, ecuación (7a), (Sec. 5.1), da la salida USSB

$$x_c(t) = 2 \operatorname{Re}[y_{LP}(t)e^{j\omega_c t}] = \tfrac{1}{2}A_c[x(t)\cos\omega_c t - \hat{x}(t)\operatorname{sen}\omega_c t]$$

Por un análisis similar, el caso de la LSSB difiere sólo en el signo del segundo término.

Entre paréntesis, se hace mención a que la ecuación (12) es proporcional a la *señal analítica*, definida por

$$x_{an}(t) \triangleq \tfrac{1}{2}[x(t) + j\hat{x}(t)] \qquad (13a)$$

Así, suponiendo que $x(t)$ es real, las partes real e imaginaria de $x_{an}(t)$ son un par transformado de Hilbert. Interesante aunque probable, la significación real descansa en el espectro, como puede comprobarlo el lector,

$$X_{an}(f) = \begin{cases} X(f) & f > 0 \\ 0 & f < 0 \end{cases} \qquad (13b)$$

así, $x_{an}^*(t)$ es precisamente la porción de frecuencia positiva de $X(f)$. Por otra parte, la transformada de Fourier de $x_{an}(t)$ es la porción de la frecuencia negativa de $X(f)$. El espectro de la SSB se forma de estas dos porciones trasladadas en frecuencia.

EJERCICIO 5.6 Confírmese la cancelación de banda lateral y la adición implicada en la ecuación (3) tomando su transformada de Fourier y empleando el hecho de que $\mathscr{F}[\hat{x}(t)] = -j \operatorname{sgn} f X(f)$.

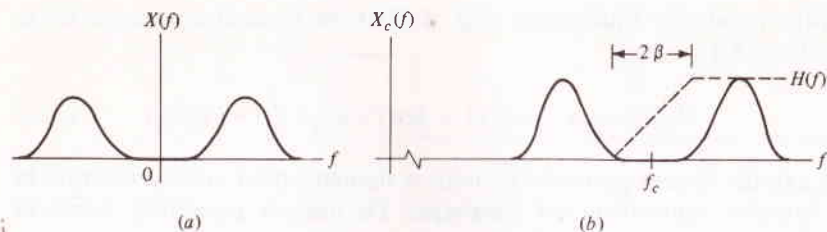



FIGURA 5.23

Moduladores de banda lateral única

La eficiencia inherente a la banda lateral única la hace bastante deseable, de manera particular cuando el *mantener el ancho de banda* constituye un factor importante. No obstante, la existencia de ciertos problemas de instrumentación se constituyen en variadas desventajas, siendo una de ellas el filtro de banda lateral incluido en la figura 5.19*a*.

La característica de corte agudo para $H(f)$ no se puede sintetizar en forma exacta, tanto que se debe o atenuar una porción de la banda lateral deseada o permitir el paso de una porción de la banda lateral no deseada. (Efectuar ambos casos equivale a modular la banda lateral residual.) Por fortuna, muchas señales moduladoras de interés práctico, tienen poco o ningún contenido de baja frecuencia, y sus espectros presentan unos "vacíos" en su frecuencia cero (Fig. 5.23*a*). Tales espectros son típicos en señales de audio (voz y música), por ejemplo. Después de la traslación hecha por el modulador balanceado, el vacío en la frecuencia cero, aparece como un espacio vacante centrado alrededor de la frecuencia portadora dentro de la cual se puede conformar la *región de transición* de un filtro de banda lateral práctico (Fig. 5.23*b*).

Como regla sencilla, diremos que el ancho 2β de la región de transición no puede ser mucho menor del 1% de la frecuencia de corte nominal, esto es, $f_{co} < 200\beta$. Puesto que 2β está limitado por el ancho de banda del vacío espectral y f_{co} debe ser igual a f_c, no es posible obtener una frecuencia portadora suficientemente alta con un espectro de mensaje dado. Para estos casos se puede realizar el proceso de modulación en dos o más pasos para superar la limitación, (véase el problema 5.18).

La ecuación (3) sugiere un medio alterno de generación de la SSB, interpretándose como dos ondas DSB con portadora en cuadratura moduladas por $x(t)$ y por $\hat{x}(t)$. Contándose con $x(t)$ y la portadora $\cos \omega_c t$, dos moduladores balanceados y un par de desplazadores de fase dispuestos (Fig. 5.24), se generará modulación en banda lateral superior (USSB) o inferior (LSSB), dependiendo de que se sumen o resten sus salidas. Esta técnica, conocida como *método de corrimiento de fase*, omite la necesidad de filtros de banda lateral. Sin embargo, el diseño de los circuitos de corrimiento de fase no es trivial, y las imperfecciones generalmente originan distorsiones en las componentes de baja frecuencia. Así, el sistema funciona mejor con espectro de mensaje del tipo analizado con anterioridad. (En el problema 5.19 se describe otro sistema de corrimiento de fase.)

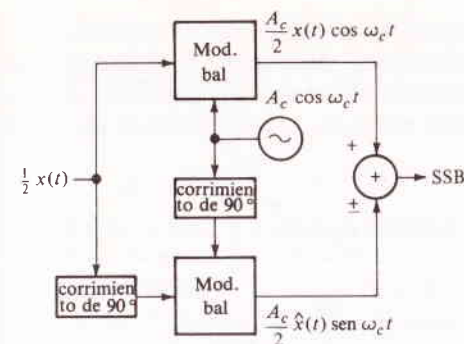



FIGURA 5.24
Modulador de SSB de corrimiento de fase.

Modulación de banda lateral residual (VSB)

Considérese una señal moduladora de ancho de banda muy grande, con un importante contenido de baja frecuencia. Los ejemplos principales son video en televisión, facsímil y señales de datos de alta velocidad. Mantener el ancho de banda es la razón del empleo de la banda lateral única, pero en los sistemas prácticos donde ella se emplea tienen una respuesta pobre en baja frecuencia. Por otra parte, la DSB funciona bastante bien para frecuencias de mensaje, pero el ancho de banda de transmisión es dos veces el de la SSB. Es obvio que se desea un esquema de modulación de compromiso; ese compromiso es la banda lateral residual (VSB).

Se llega a la banda lateral residual por medio del filtraje de la DSB (o AM) en una modalidad que una banda lateral pase casi completamente, mientras que sólo se incluye un trozo o *vestigio* de la otra banda lateral. La clave para llegar a la VSB es el filtro de banda lateral, una función de transferencia típica, representada en la figura 5.25*a*. En tanto que la forma exacta de la respuesta no es crucial, debe tener simetría impar en relación con la frecuencia portadora y una respuesta relativa de ½ en ese punto. Por lo que, tomando el caso de la banda lateral superior,

$$H(f) = u(f - f_c) - H_\beta(f - f_c) + u(-(f + f_c)) - H_\beta(-(f + f_c)) \qquad (14a)$$

donde

$$H_\beta(-f) = -H_\beta(f) \qquad \text{y} \qquad H_\beta(f) = 0 \qquad |f| > \beta \qquad (14b)$$

como se muestra en la figura 5.25*b*.

El filtro de banda lateral para VSB es así un filtro de banda lateral *práctico* con ancho de transición 2β, con lo que un modulador† de VSB toma la forma de la figura 5.19*a*. (Si no se desea suprimir la portadora, el modulador balanceado se sustituye por un modulador de AM.) Como el ancho de la banda lateral parcial es la mitad del ancho de transición del filtro, el ancho de banda de transmisión es

$$B_T = W + \beta \approx W \qquad (15)$$

Los espectros de VSB y de SSB son muy similares, de manera particular cuando $\beta \ll W$, lo cual es verdad a menudo. Lo mismo respecto de las similitudes que

† En la práctica, la simetría del filtro que se requiere se efectúa principalmente en el receptor.

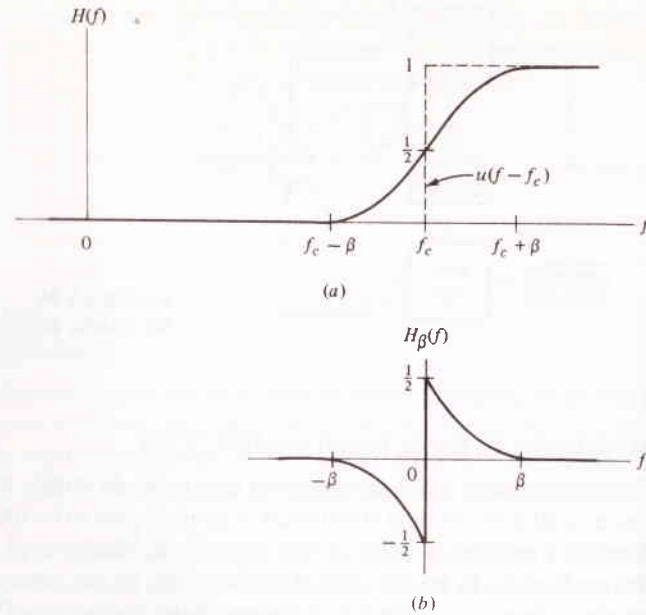



FIGURA 5.25
Características del filtro de VSB.

existen en el dominio del tiempo, pudiéndose escribir $x_c(t)$ en una forma paralela a la de la ecuación (3), o sea,

$$x_c(t) = \tfrac{1}{2}A_c\{x(t)\cos\omega_c t - [\hat{x}(t) + x_\beta(t)]\,\mathrm{sen}\,\omega_c t\} \qquad (16)$$

donde la componente en cuadratura es $\hat{x}(t)$ más

$$x_\beta(t) \triangleq j2\int_{-\beta}^{\beta} H_\beta(f)X(f)e^{j\omega t}\,df$$

Si $\beta \ll W$, la VSB se aproxima a la SSB y $x_\beta(t) \approx 0$; a la inversa, para β grande, la VSB se aproxima a la DSB y $\hat{x}(t) + x_\beta(t) \approx 0$. La potencia transmitida S_T no es muy fácil de determinar en forma exacta, pues queda limitada por

$$\tfrac{1}{4}\overline{x^2}A_c^2 \le S_T \le \tfrac{1}{2}\overline{x^2}A_c^2 \qquad (17)$$

dependiendo del ancho residual β.

Banda lateral única compatible ★

En ocasiones se desea combinar la modulación de envolvente de AM con la conservación del ancho de la banda lateral suprimida. El objeto es permitir la detección de la envolvente sin *pagar el precio* de un ancho de banda de transmisión excesivo; de ahí el nombre de *banda lateral única compatible.*

Aunque una modulación de envolvente perfecta necesita bandas laterales simétricas, se puede aproximar por medio de la modulación de banda lateral suprimida con portadora no suprimida o reinsertada. Agregando un término portadora e índice de modulación a la ecuación (16), la onda modulada resultante es

$$x_c(t) = A_c\{[1 + mx(t)]\cos\omega_c t - m\zeta(t)\,\mathrm{sen}\,\omega_c t\} \qquad (18)$$

donde $\zeta(t) = \hat{x}(t) + x_\beta(t)$ es el término en cuadratura que cancela una banda lateral en todo o en parte. Si $\zeta(t) = 0$, la ecuación (18) se reduce a la *modulación en amplitud* convencional; si $\zeta(t) = \mp\hat{x}$, se tiene una SSB más portadora (SSB + C); para la VSB + C, $\zeta(t)$ se toma sobre un valor intermedio.

La envolvente de $x_c(t)$ se encuentra en la modalidad usual que es

$$\begin{aligned} R(t) &= A_c\{[1 + mx(t)]^2 + [m\zeta(t)]^2\}^{1/2} \\ &= A_c[1 + mx(t)]\left\{1 + \left[\frac{m\zeta(t)}{1 + mx(t)}\right]^2\right\}^{1/2} \end{aligned} \qquad (19)$$

la cual es una envolvente de AM distorsionada. Sin embargo, para $|m\zeta(t)| \ll 2$, la envolvente es, en forma aproximada

$$R(t) \approx A_c[1 + mx(t)]$$

como se desea. La clave para la banda lateral única compatible es, por lo tanto, mantener pequeña la componente en cuadratura.

Si una banda lateral se suprime *en forma total*, es decir, SSB + C, no se puede despreciar a $\zeta(t)$; la modulación de envolvente efectiva necesita entonces que $m \ll 1$. Bajo esta condición se desperdicia una cantidad sustancial de potencia en la portadora, mucho más que en el caso de la AM. En consecuencia, existe un marcado interés en otros esquemas para la generación de banda lateral única compatible.†

De manera alterna, si se puede tolerar una banda lateral *residual* y su ancho no es muy pequeño, $|\zeta(t)|$ es pequeño comparado con $|x(t)|$, la mayor parte del tiempo. Se puede emplear entonces un índice de modulación mayor sin excesiva distorsión de envolvente, y reducir con ello la potencia de portadora relativa. En efecto, se demuestra con facilidad, en la ecuación (18), que

$$S_T = [1 + m^2(\overline{x^2} + \overline{\zeta^2})](A_c^2/2) \approx (1 + m^2\overline{x^2})P_c \qquad (20)$$

en esencia, lo mismo que en la AM.

Se debe determinar el necesario ancho de banda lateral residual por medio de estudios empíricos de la envolvente con material de programa típico. En el caso de las señales de video en televisión, las cuales se transmiten como VSB + C, la distorsión puede ser bastante grande sin menoscabo de la calidad de la imagen, permitiendo un ancho de banda de transmisión sólo 30% mayor que el ancho de banda del mensaje. Véase el Apéndice C para mayor información de los sistemas de TV.

† En Schwartz, Bennett y Stein (1966, Cap. 4) se dan un resumen y referencias.

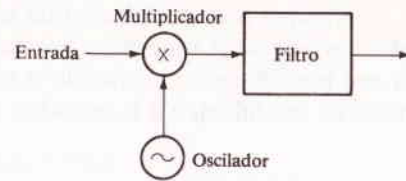



FIGURA 5.26
Convertidor de frecuencia.

5.5 CONVERSIÓN DE FRECUENCIA, DETECCIÓN Y RECEPTORES

La modulación lineal es principalmente una traslación de frecuencia directa del espectro del mensaje. La *demodulación* o *detección*, es el proceso, en el receptor, por medio del cual se recupera el mensaje, de la onda modulada. Por lo que, para la modulación lineal en general, el proceso de detección es en forma básica uno de los casos de *traslación de frecuencia a valores menores*.

Con relación a las figuras 5.8 y 5.19*b*, por ejemplo, se puede ver que si los espectros se desplazan hacia abajo en frecuencia en f_c unidades (hacia arriba en f_c unidades para las componentes de frecuencia negativa), se reproduce el espectro del mensaje original, más una posible componente de CD correspondiente a la portadora trasladada. Se insta al lector a efectuar esta traslación en forma gráfica, por sí mismo. Es particularmente instructivo en caso de la LSSB.

La traslación de frecuencia, o *conversión*, se emplea también para desplazar una señal modulada a otra nueva frecuencia portadora (arriba o abajo) para amplificación o para algún otro procesamiento. Así, la traslación es una operación fundamental de los sistemas de modulación lineal e incluye la modulación y la detección como casos especiales. Antes de examinar los detectores, debe hacerse un breve examen al proceso general de la conversión de frecuencia.

Conversión de frecuencia

La conversión se efectúa, al menos en forma analítica, al multiplicar por una sinusoide. Considérese, por ejemplo, la onda de DSB $x(t)\cos\omega_1 t$. Multiplicando por $\cos\omega_2 t$, se obtiene

$$x(t)\cos\omega_1 t\cos\omega_2 t = \tfrac{1}{2}x(t)\cos(\omega_1+\omega_2)t + \tfrac{1}{2}x(t)\cos(\omega_1-\omega_2)t \qquad (1)$$

El producto está compuesto de las *frecuencias suma y diferencia*, $f_1 + f_2$ y $|f_1 - f_2|$, cada una modulada por $x(t)$. [Se puede escribir $|f_1 - f_2|$ para claridad, puesto que $\cos(\omega_2 - \omega_1)t = \cos(\omega_1 - \omega_2)t$. Suponiendo que $f_2 \neq f_1$, la multiplicación ha trasladado el espectro de la señal a *dos* nuevas frecuencias portadoras. Con un filtraje apropiado, la señal se convierte a un valor mayor o a uno menor en frecuencia. Los dispositivos que realizan esta operación se denominan *convertidores de frecuencia o mezcladores*. La operación en sí es designa como *heterodinación* o *mezcla*.

La figura 5.26 muestra el diagrama de un convertidor de frecuencia generalizado. Su análisis se apega en forma directa al teorema de la modulación. El

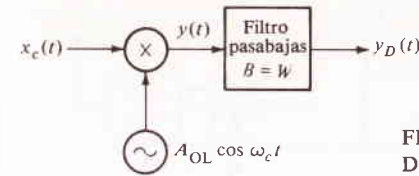



FIGURA 5.27
Detección sincrónica (de producto).

multiplicador se construye por lo general empleando dispositivos no lineales o de conmutación, similares a los moduladores. Otras aplicaciones de los convertidores comprenden osciladores de frecuencia de batido, divisores de frecuencia regenerativos, perturbadores mezcladores de voz y analizadores de espectro.

EJERCICIO 5.7 Bosquéjese el espectro de la ecuación (1) para $f_2 < f_1, f_2 = f_1$ y $f_2 > f_1$, tomando a $X(f)$ como en la figura 5.6.

Detección sincrónica

Todos los tipos de modulación lineal se pueden detectar por medio del demodulador *producto* de la figura 5.27. La señal de entrada primero se multiplica con una sinusoide que se genera en forma local y luego se pasa a través de un filtro pasabajas, siendo el ancho de banda el mismo que el del mensaje, o sea W, o algo mayor. Se supone que el oscilador local (OL) está *sincronizado con exactitud* con la portadora, tanto en fase como en frecuencia, a lo cual se le da el nombre de *detección sincrónica*.†

Para propósitos de análisis, escríbase la señal de entrada en la forma generalizada

$$x_c(t) = [K_c + K_m x(t)]\cos\omega_c t - K_m\zeta(t)\operatorname{sen}\omega_c t \qquad (2)$$

lo cual puede representar a algún tipo de modulación lineal, con una apropiada identificación de K_c, K_m y $\zeta(t)$ —por ejemplo, $K_c = 0$ para portadora suprimida, $(t) = 0$ para doble banda lateral, etc. La entrada del filtro es así el producto

$$x_c(t)A_{OL}\cos\omega_c t = \frac{A_{OL}}{2}\{[K_c + K_m x(t)] + [K_c + K_m x(t)]\cos 2\omega_c t - K_m\zeta(t)\operatorname{sen}2\omega_c t\}$$

Como $f_c > W$, el filtro pasabajas rechaza los términos de frecuencia doble, permitiendo sólo el paso del primer término

$$y_D(t) = K_D[K_c + K_m x(t)] \qquad (3)$$

donde K_D es la constante de detección. La componente de CD, $K_D K_c$ corresponde a la portadora trasladada si no está suprimida en la onda modulada. Se puede remover

† También se le designa como *detección coherente*, porque deben ser coherentes la onda portadora y la salida del oscilador local.

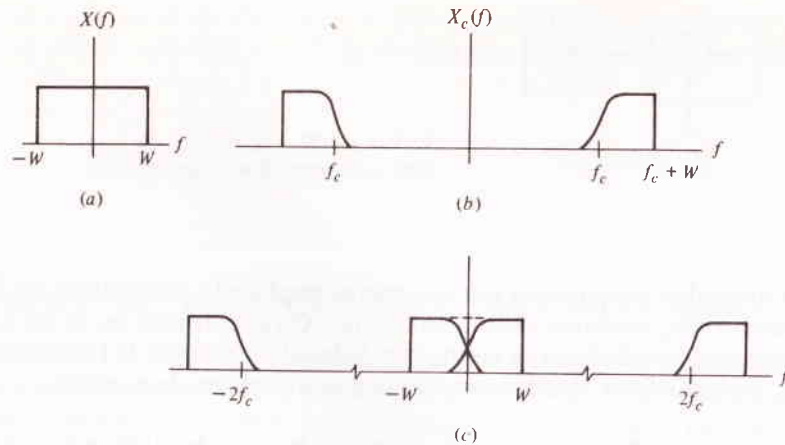



FIGURA 5.28
Espectros de VSB. (*a*) mensaje; (*b*) señal modulada; (*c*) señal de frecuencia trasladada antes del filtraje pasabandas.

de la salida por medio de un capacitor de bloqueo o de un transformador; sin embargo, asimismo se removerá algún término de CD en $x(t)$. Con esta restricción menor, se puede decir que el mensaje ha sido recuperado en forma total de $x_c(t)$.

Aunque correctos en grado sumo, los desarrollos anteriores fallan al extraer lo que existe en la demodulación de la VSB. Esto se ve mejor en el dominio de la frecuencia, considerando el espectro del mensaje constante sobre W, figura 5.28*a*, donde el espectro modulado adopta la forma de la figura 5.28*b*. El espectro trasladado hacia abajo en la entrada del filtro será entonces como el mostrado en la figura 5.28*c*. De nuevo, los términos de alta frecuencia se eliminan por medio del filtraje, en tanto que las bandas laterales convertidas a valores menores, se traslapan alrededor del valor cero de frecuencia. Recordando la propiedad de simetría del filtro residual, se encuentra que la porción removida de la banda lateral superior se restituye en forma completa por el vestigio correspondiente de la banda lateral inferior, reconstruyéndose así $X(f)$ a la salida, y la señal detectada resulta proporcional a $x(t)$.

En teoría, la demodulación producto raya en lo trivial; en la práctica, puede ser bastante artificiosa. Lo difícil del problema es la *sincronización* —(sincronizar un oscilador con una sinusoide que no está siempre presente en la señal que llega en caso de estar suprimida la portadora). Para facilitar el asunto, los sistemas de portadora suprimida pueden tener una cantidad pequeña de portadora reinsertada en $x_c(t)$ en el transmisor. Esta *portadora piloto* se extrae en el receptor por medio de un filtro pasabanda de banda angosta, se amplifica, y se emplea en lugar de un oscilador local. Al sistema mostrado en la figura 5.29, se le designa *detección homodina*. (En realidad, la portadora piloto amplificada se emplea más a menudo para sincronizar un oscilador separado que en forma directa.)

Es posible una variedad de otras técnicas para sincronización, incluyendo el empleo de osciladores controlados a cristal, altamente estables, en transmisor y re-

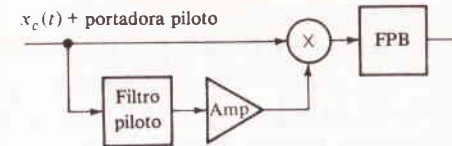



FIGURA 5.29
Detección homodina.

ceptor, estando los osciladores sincronizados en forma periódica más que en una base continua. No obstante, se debe esperar algún grado de asincronismo en los detectores sincrónicos. Es, por lo tanto, importante investigar los efectos del corrimiento de fase y frecuencia en diferentes aplicaciones. Esto se hará para la DSB y la SSB en términos de la modulación de tono.

Sea $\cos(\omega_c t + \omega' t + \phi')$, la onda del oscilador local, donde ω' y ϕ' representan la frecuencia de corrimiento lento y los errores de fase comparados con la portadora. Para doble banda lateral con modulación de tono la señal detectada resulta ser

$$y_D(t) = K_D \cos \omega_m t \cos(\omega' t + \phi') \tag{4}$$

$$= \begin{cases} \dfrac{K_D}{2}[\cos(\omega_m + \omega')t + \cos(\omega_m - \omega')t] & \phi' = 0 \\ K_D \cos \omega_m t \cos \phi' & \omega' = 0 \end{cases}$$

De manera similar, para banda lateral única, donde $x_c(t) = \cos(\omega_c \pm \omega_m)t$,

$$y_D(t) = K_D \cos[\omega_m t \pm (\omega' t + \phi')] \tag{5}$$

$$= \begin{cases} K_D \cos(\omega_m \pm \omega')t & \phi' = 0 \\ K_D \cos(\omega_m t \pm \phi') & \omega' = 0 \end{cases}$$

Es obvio, que tanto en DSB como en SSB, un corrimiento de frecuencia que no es pequeño comparado con W alterará en forma sustancial el tono detectado. El efecto es más severo en DSB, puesto que se produce un *par* de tonos, $f_m + f'$ y $f_m - f'$. Si $f' \ll f_m$, se escuchará un sonido melodioso o la nota de batido que se escucha cuando se tocan al unísono dos instrumentos musicales, estando ligeramente fuera de tono. Aunque sólo se produce un tono con SSB, éste puede también provocar disturbios, en forma particular para transmisión de música. Como ilustración, la triada de cuerdas mayor consiste de tres notas cuyas frecuencias están relacionadas con los enteros 4, 5 y 6. El error de frecuencia en la detección desplaza cada nota en la misma cantidad absoluta, destruyendo la relación armónica y dando a la música un sabor oriental. (Nótese que el efecto *no* es como el de la música grabada que se reproduce a una velocidad incorrecta, en cuyo caso se conservan las relaciones de frecuencia.) Para transmisión de voz, las pruebas con oyentes subjetivos han demostrado que son tolerables corrimientos de frecuencia menores que ± 10 Hz; de otra manera, cualquier sonido se escuchará más bien como la "voz" del pato Donald.

En la misma forma, en corrimiento de fase, de nuevo la DSB es más sensitiva, porque si $\phi' = \pm 90°$ (oscilador local y portadora en cuadratura), la señal detectada

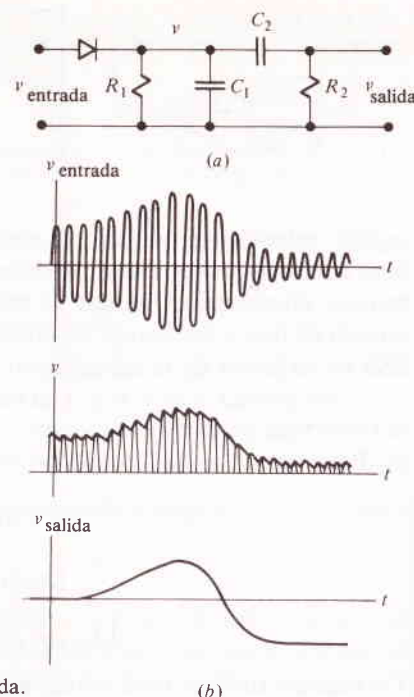



FIGURA 5.30
Detección de envolvente. (*a*) circuito; (*b*) formas de onda.

desaparece por completo. Con pequeñas variaciones de ϕ', se obtiene un efecto de *desvanecimiento* aparente. El corrimiento de fase en la SSB aparece como *distorsión por retardo*, siendo el caso extremo cuando $\phi = \pm 90°$, para lo cual la señal demodulada es $\hat{x}(t)$. Sin embargo, como se recalcó antes, el oído humano puede tolerar una distorsión por retardo grande por lo que el corrimiento de fase no es tan serio en sistemas de SSB para señales de voz.

Resumiendo, los requisitos de sincronización de fase y frecuencia son más modestos para transmisión de voz por medio de la SSB. Pero en sistemas de datos, de facsímil y de video con portadora suprimida, se necesita una sincronización cuidadosa. Principalmente por esa razón, esta portadora no se suprime en transmisión de televisión.

Detección de envolvente

Muy poco se ha dicho hasta aquí acerca de la demodulación sincrónica de AM, por la sencilla razón de que casi nunca se emplea. En realidad los detectores funcionan en AM, pero en forma de *detectores de envolvente*, los cuales son mucho más sencillos. Como la envolvente de una onda de AM tiene la misma forma del mensaje, independiente de la frecuencia y fase de la portadora, se puede llevar a cabo la demodulación extrayendo la envolvente sin preocuparse por la sincronización.

Un detector de envolvente simplificado y sus formas de onda, se pueden ver en la figura 5.30, donde se ha supuesto que el diodo es lineal. En ausencia del resto del

circuito, el voltaje v sería precisamente la versión rectificada de media onda de la entrada $v_{entrada}$. Pero R_1C_1 actúan como un filtro pasabajas, que responde sólo a las variaciones en los picos de $v_{entrada}$. Esto supone la constante de tiempo grande comparada con $1/f_c$ y pequeña en comparación con el tiempo de variación del mensaje $1/W$. Así, como se apuntó antes, se necesita que $f_c \gg W$ para que la envolvente quede definida en forma clara. Bajo estas condiciones, C_1 se descarga sólo un poco entre picos de la portadora, y v será en forma aproximada la envolvente de $v_{entrada}$. Un filtraje más complejo ocasionará una considerable mejoría si fuera necesario. Por último, R_2C_2 actúan como un medio de bloqueo de la CD para remover la polarización de la componente no modulada de la portadora.

El voltaje v también se puede filtrar para remover las variaciones de la envolvente y producir un voltaje de CD proporcional a la amplitud de la portadora. Este voltaje, a su vez se realimenta a los pasos precedentes del receptor para proporcionar *control automático de volumen* (AVC) y compensar el desvanecimiento.

A despecho del elemento no lineal, el circuito de la figura 5.30 se designa como *detector lineal de envolvente*; la salida es proporcional en forma lineal a la envolvente de entrada. Se pueden usar también diodos de ley de potencia, pero entonces v incluirá términos de la forma $v^2_{entrada}$, $v^3_{entrada}$, etc., y puede haber una apreciable distorsión por segunda armónica a menos que $m \ll 1$. En efecto, el diseño de detectores lineales de envolvente tiene varios problemas interesantes que no aparecen en forma inmediata en el presente análisis†.

La detección de envolvente es apropiada siempre y cuando la onda en cuestión tenga una modulación de envolvente apropiada, por ejemplo, el video de televisión transmitido por medio de banda lateral residual más portadora. Además, en algunos sistemas de portadora suprimida se emplean más los detectores de envolvente que los demoduladores de producto, insertándose un término grande de portadora adelante del detector para *reconstruir* la envolvente. (Esto se hace extensivo a la SSB y a la DSB si la portadora local es suficientemente grande.) Pero el procedimiento no evita la necesidad de sincronismo, puesto que la reconstrucción de la envolvente requiere que la portadora agregada esté bien sincronizada.

Un punto final. El capacitor de bloqueo de la CD de la figura 5.30 origina que el detector tenga una respuesta pobre a las componentes de baja frecuencia del mensaje. Por lo tanto, la detección de envolvente puede no ser satisfactoria para señales con un importante contenido de CD y términos que varíen lentamente, a menos que se tengan pasos adicionales.

EJERCICIO 5.8 La figura 5.31 es un sistema de *reconstrucción de envolvente* para la demodulación de una modulación de portadora suprimida; nótese que la salida del oscilador local se *agrega* a $x_c(t)$ por lo que no se requiere de un multiplicador. Tomando $x_c(t) = A_c[x(t) \cos \omega_c t - \zeta(t) \operatorname{sen} \omega_c t]$ para representar DSB, SSB o VSB, y tomando la señal del oscilador local como $A_{OL} \cos [\omega_c t + \theta(t)]$ donde $\theta(t) =$

† Véase, por ejemplo, Clarke y Hess (1971, Cap. 10).

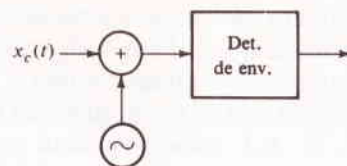



FIGURA 5.31
Sistema para detectar la modulación de portadora suprimida por medio de la reconstrucción de envolvente.

$\omega' t + \phi'$ se ajusta para posible asincronismo, demuéstrese que la envolvente en la entrada del detector de envolvente es

$$R(t) = A_{OL}\left\{1 + \frac{2A_c}{A_{OL}}[x(t)\cos\theta(t) + \zeta(t)\operatorname{sen}\theta(t)] + \left(\frac{A_c}{A_{OL}}\right)^2 [x^2(t) + \zeta^2(t)]\right\}^{1/2} \qquad (6)$$

Luego demuéstrese que puede recuperarse $x(t)$ sin distorsión de cuadratura siempre y cuando $A_{OL} \gg A_c$ y $\theta(t) \approx 0$.

Receptores

Todo lo que en realidad se requiere en un receptor es algún mecanismo de sintonía, un demodulador y amplificadores. Con una señal de suficiente intensidad, incluso se puede omitir los amplificadores, lo cual se comprueba con el histórico receptor a cristal. Muchos radiorreceptores, sin embargo, son del tipo *superheterodino* más complejo, cuyo diagrama a bloques se aprecia en la figura 5.32, completo incluyendo AVC.

Existen tres tipos de amplificadores en un superheterodino: el amplificador de *radiofrecuencia* (RF), el cual se sintoniza a la frecuencia portadora deseada; el amplificador de *frecuencia intermedia* (FI), de sintonía fija y que da lugar a la mayor parte de la ganancia y selectividad; y el amplificador *de audiofrecuencia* (AF), el cual se encuentra a continuación del detector y eleva la potencia al nivel requerido por el altoparlante. El mezclador es un convertidor de frecuencia que traslada la salida de RF a la banda de FI convirtiéndola f_c a la f_{FI}. El oscilador local (OL) proporciona la frecuencia de mezcla y se ajusta en paralelo con la etapa de RF.

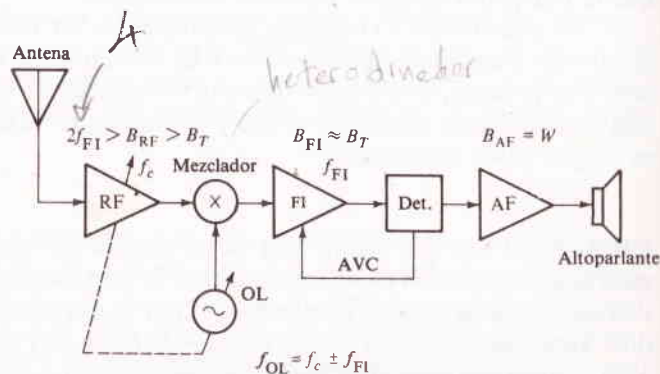



FIGURA 5.32
Receptor superheterodino.

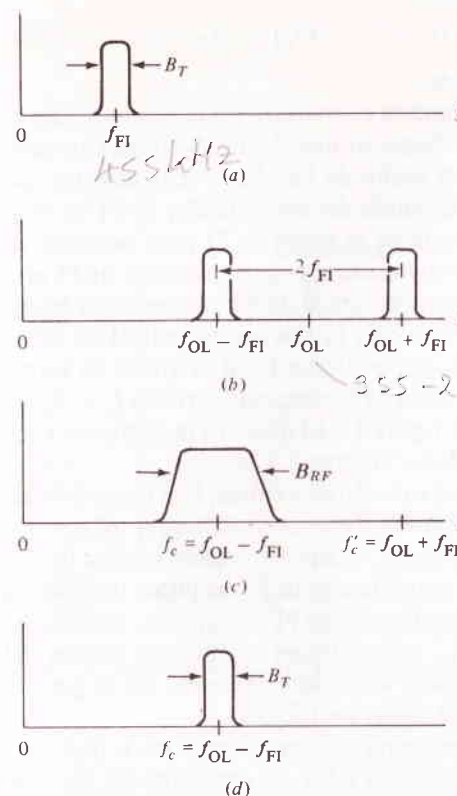



FIGURA 5.33
Características de respuesta en frecuencia de un superheterodino: (*a*) El amplificador de FI solo; (*b*) el mezclador y amplificador de FI; (*c*) el amplificador de RF solo; (*d*) el receptor completo.

Se comprenden mejor las operaciones en un superheterodino considerando la amplitud relativa de la salida de FI como una función de la frecuencia de entrada en varios puntos de receptor. Si se aplica una sinusoide de frecuencia variable y de amplitud constante a la entrada del amplificador de FI, la respuesta es justamente la relación de amplitudes del amplificador mismo (Fig. 5.33*a*). Puesto que el amplificador de FI debe permitir el paso de la señal modulada a la frecuencia portadora trasladada, su ancho de banda debe ser igual o mayor que B_T. Coniderando el mezclador más el oscilador local solos, hay *dos* frecuencias de entrada que serán la f_{FI} en la salida, o sean $f_{OL} \pm f_{FI}$. Así, la respuesta del amplificador de FI y del mezclador será como se muestra en la figura 5.33*b* para el caso común donde $f_{FI} < f_{OL}$. Nótese que si se está tratando de recibir una estación en $f_c = f_{OL} - f_{FI}$, sólo se captará $f_c = f_{OL} + f_{FI}$, la cual se conoce como *frecuencia imagen*. Para una posición dada del oscilador local, la portadora y las frecuencias imágenes están relacionadas por

$$|f_c - f'_c| = 2f_{FI} \qquad f_{FI} < f_{OL} \qquad (7a)$$

o

$$|f_c - f_c'| = 2f_{OL} \qquad f_{FI} > f_{OL} \qquad (7b)$$

El último caso es bastante raro en la práctica.

El objeto de la etapa de RF, cuya respuesta se muestra en la figura 5.33*c*, es rechazar la frecuencia imagen antes de que llegue al mezclador. Se puede apreciar que el amplificador de RF necesita tener un ancho de banda no más angosto que alrededor de $2f_{FI}$. Por otra parte, el ancho de banda del amplificador de FI no debe ser mayor que B_T, porque éste está por encima de la etapa de FI para rechazar las portadoras en la vecindad inmediata de la señal deseada, es decir, la etapa de FI proporciona *selectividad de canal adyacente*, mientras que la de RF proporciona *rechazo de canal de imagen*. En operación convencional, la frecuencia central de RF se sintoniza a la portadora deseada mientras que el oscilador local se ajusta en forma simultánea a $f_{OL} = f_c + f_{FI}$ para que la diferencia de frecuencias correcta $f_c - f_{OL} = f_{FI}$ se obtenga en la salida del mezclador. La figura 5.33*d* muestra la respuesta total del receptor obtenida de la multiplicación de las figuras 5.33*b* y *c*.

Hasta aquí se ha considerado sólo un superheterodino ideal. Los receptores reales presentarán *respuestas espurias* a varias otras frecuencias debido a señales que entran al receptor o a no linealidades. Por ejemplo, cuando una señal intensa de frecuencia cercana a ½ f_{FI} llega a la entrada del amplificador de FI, se puede producir su segunda armónica si la primera etapa del amplificador de FI no es lineal. Esta segunda armónica, que es en forma aproximada f_{FI}, será entonces amplificada por las etapas siguientes y aparecerá a la salida del detector como interferencia, por lo general como un tono de audio de alta frecuencia, o como un silbido.

Otros tipos de receptores que se emplean para diferentes propósitos incluyen: el receptor *heterodino*, en el cual se omite la etapa de RF y, en consecuencia, las imágenes constituyen un problema potencial; el receptor de *radiofrecuencia sintonizada* (TRF), en el cual la detección se efectúa en seguida de la etapa de RF; y el receptor de *doble conversión*, que se emplea comúnmente para onda corta de alta calidad en AM y en sistemas de banda lateral única. Como sugiere el nombre, los receptores de doble conversión tienen dos mezcladores y dos amplificadores de FI. El primer FI se sintoniza a frecuencia fija sobre o inmediata por debajo de la banda de la frecuencia portadora deseada. El valor grande de f_{FI_1} aumenta el espaciamiento $2f_{FI_1}$ entre una portadora y su imagen, mejorando el rechazo de imagen de la etapa de RF. El segundo FI tiene una frecuencia central relativamente baja para permitir una discriminación mucho más aguda contra canales adyacentes lo que no es posible de otra manera.

Ejemplo 5.3 Receptor superheterodino de AM

La distribución de frecuencias portadoras en AM comercial va de los 540 a los 1 600 kHz con 10 kHz de espaciamiento, y el ancho de banda del mensaje es de unos 5 kHz. Muchas radios de AM son superheterodinos con f_{FI} = 455 kHz y $f_{OL} = f_c + f_{FI}$. Empleando estos valores, los anchos de banda del receptor son

$$10\text{ kHz} < B_{RF} < 910\text{ kHz} \qquad B_{FI} \approx 10\text{ kHz} \qquad B_{AF} \approx 5\text{ kHz}$$

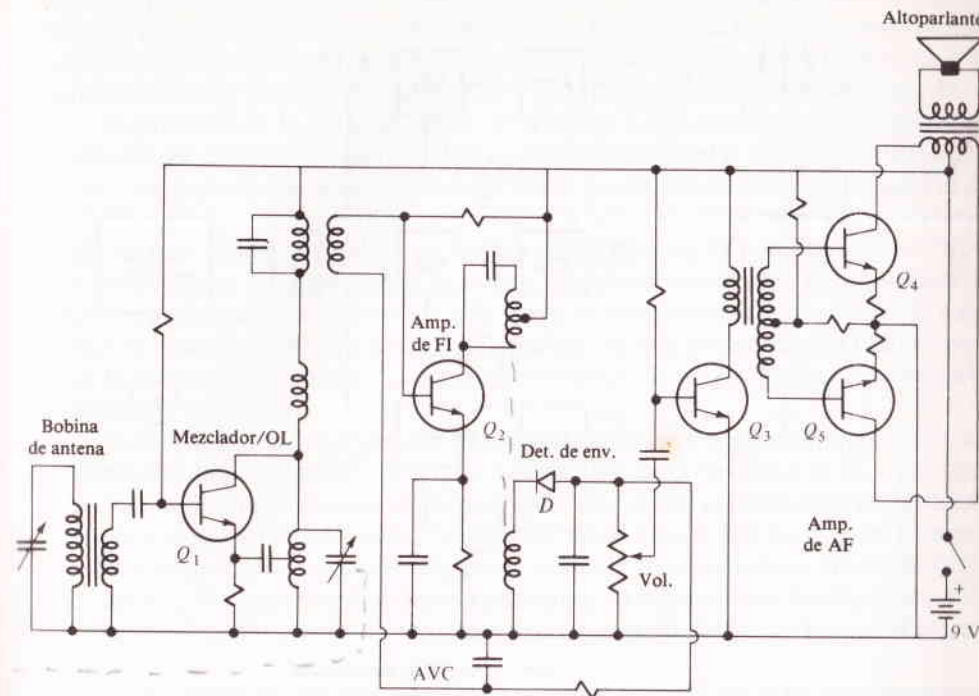



FIGURA 5.34
Diagrama esquemático de una radio superheterodino de AM.

El intervalo de sintonía de RF es, por supuesto, de 540 a 1 600 kHz, mientras que el oscilador local debe sintonizar de 955 a 2 055 kHz. La relación de sintonía del oscilador local es, por lo tanto, *grosso modo*, 2:1. Por otra parte, si f_{OL} se toma como $f_c - f_{FI}$, entonces el intervalo de sintonía sería de 85 a 1 145 kHz; esto corresponde a una relación de 13:1 y sería mucho más difícil llevarla a cabo con capacitores variables, etc.

La figura 5.34 muestra el diagrama esquemático de una radio de AM típica, con alimentación a baterías y económica. El transistor Q_1 sirve como oscilador local y mezclador; el circuito sintonizado conectado a su emisor está acoplado en forma magnética a la bobina en el circuito del colector, dando lugar a la oscilación a una frecuencia que determina el capacitor variable, el cual a su vez, está acoplado en forma mecánica al capacitor de sintonía en la bobina de antena de tal manera que se logra el rastreo simultáneo de f_c y f_{OL}. No existe amplificación, pero el circuito de la bobina de antena produce rechazo de imagen y entrega la frecuencia portadora deseada a la base de Q_1. La no linealidad produce las frecuencias $f_c + f_{OL}$ y $f_c - f_{OL}$ en la corriente del colector y $f_{FI} = f_c - f_{OL}$ se selecciona por medio del circuito de colector sintonizado y acoplado al amplificador de FI Q_2. La salida de FI impulsa al diodo D, por medio de acoplamiento magnético, y la señal de envolvente se aplica a la primera etapa del amplificador de audio Q_3 a través del control de volumen. Ade-

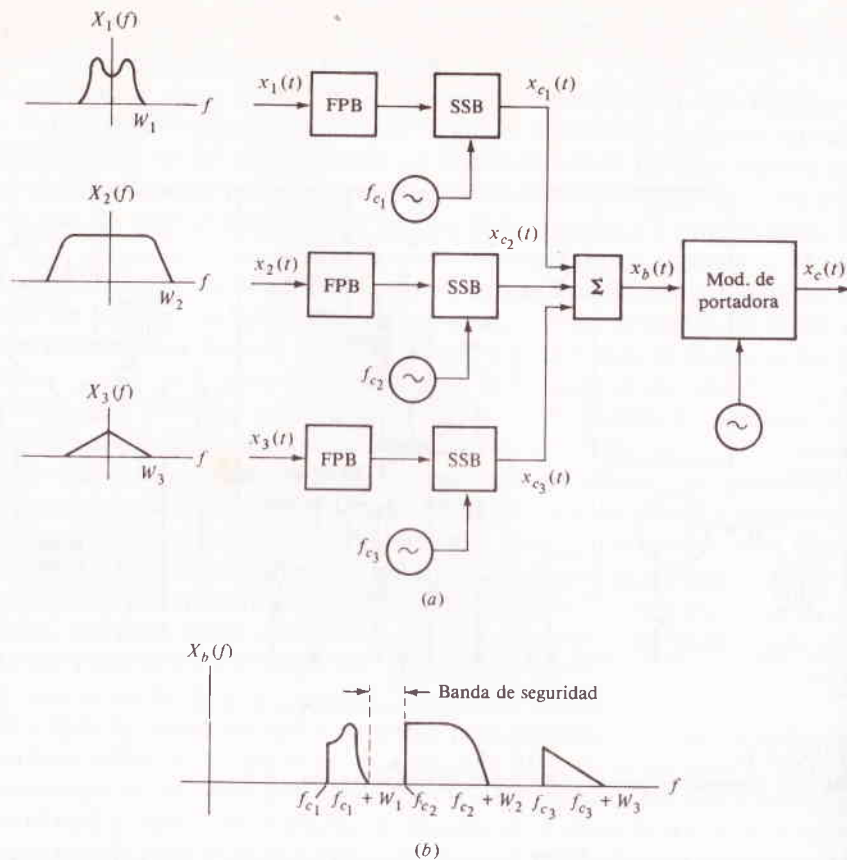



FIGURA 5.35
Transmisor de multicanalización por división en frecuencia. (*a*) Diagrama a bloques; (*b*) espectro de banda base.

más, el voltaje promedio de envolvente se realimenta a Q_2 para modificar la polarización de base para proporcionar control automático de volumen. Los transistores Q_4 y Q_5 constituyen un amplificador de audio clase B que impulsa al altoparlante.

EJERCICIO 5.9 Se necesita un receptor superheterodino para cubrir f_c = 1 a 3 MHz con $f' \geq 5$ MHz y $f_{FI} \searrow f_{OL}$. Encuéntrense los valores apropiados para f_{FI} y f_{OL}. *Respuesta*: 2 MHz, 3 a 5 MHz.

5.6 MULTICANALIZACIÓN POR DIVISIÓN EN FRECUENCIA

A menudo se desea transmitir varios mensajes con facilidad de transmisión, un proceso conocido como *multicanalización* . * Las aplicaciones de la multicanalización van

* *N. del E.* Es frecuente el uso del término *multiplexado*. Aquí sólo utilizaremos *multicanalización*.

de la vital, aunque prosaica, red telefónica hasta la fascinante FM estéreo y los sistemas de telemetría por sondas espaciales. Existen dos técnicas básicas de multicanalización: la *multicanalización por división en frecuencia* (FDM) que se trata aquí, y la *multicanalización por división en tiempo* (TDM) que luego se examina (Cap. 8).

El principio de la FDM se ilustra en la figura 5.35*a*, en donde varios mensajes de entrada (se muestran tres) modulan en forma individual a las *subportadoras* f_{c_1}, f_{c_2}, etc., después de que pasan a través de filtros pasabajas para limitar los anchos de banda de mensaje. Se muestra la modulación de subportadora como SSB, y a menudo así es; pero se puede emplear cualquiera de las técnicas de modulación de CW, o una combinación de ellas. Luego se suman las señales moduladas para producir la *señal de banda base*, con espectro $X_b(f)$, como se muestra en la figura 5.35*b*. (La designación "banda base" se emplea aquí para indicar que la modulación final de portador aún no ha tenido lugar.) La función del tiempo de banda base $x_b(t)$ se deja a la imaginación del lector.

Suponiendo que se escogen en forma apropiada las frecuencias subportadoras, la operación multicanalización asigna una ranura en el dominio de la frecuencia para cada uno de los mensajes en forma modulada; de ahí el nombre de multicanalización *por división en frecuencia*. La señal de banda base puede después ser transmitida en forma directa o empleada para modular una portadora transmitida de frecuencia f_c. No existe aquí un interés particular en relación con la naturaleza de la modulación final de portadora, puesto que el espectro de la banda base es el importante.

La recuperación del mensaje o la demodulación de FDM se hace en tres pasos. (Fig. 5.36). Primero, el demodulador de portadora reproduce la señal de banda base $x_b(t)$. Luego se separan las subportadoras moduladas por medio de un banco de filtros pasabanda en paralelo, y en seguida se detectan cada uno de los mensajes.

El principal problema práctico en la FDM es la *diafonía*, la indeseable unión de un mensaje con otro. La diafonía inteligible (modulación cruzada) aparece principalmente por las no linealidades del sistema, lo cual ocasiona que una señal de

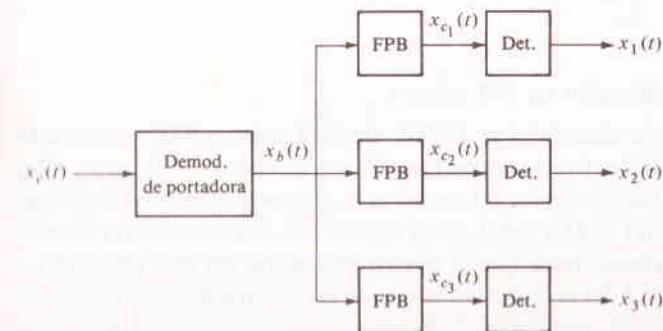



FIGURA 5.36
Receptor de multicanalización por división en frecuencia.

Tabla 5.2 JERARQUÍA DE LA MULTICANALIZACIÓN POR DIVISIÓN EN FRECUENCIA DEL SISTEMA BELL

Designación	Intervalo de frecuencia	Ancho de banda	Número de canales de voz
Grupo	60–108 kHz	48 kHz	12
Supergrupo	312–552 kHz	240 kHz	60
Grupo maestro	564–3,084 kHz	2.52 MHz	600
Grupo Jumbo	0.5–17.5 MHz	17 MHz	3,600

mensaje module en forma parcial a otra subportadora.† En la misma forma, la diafonía ininteligible ocasiona disturbios por la imperfecta separación espectral en el banco de filtros. Para reducir esta unión o acoplamiento, los espectros de mensaje modulados se separan en frecuencia por medio de bandas de seguridad en las que se pueden distribuir las regiones de transición de los filtros. Por ejemplo, la banda de seguridad entre $X_{c_1}(f)$ y $X_{c_1}(f)$ en la figura 5.35*b*, es del ancho $f_{c_2} - (f_{c_1} + W_1)$. El *ancho de banda de la banda base* neto es, por lo tanto, la suma de los anchos de banda de mensaje modulados más las bandas de seguridad.

No obstante que los conceptos relativos a la FDM son muy sencillos, los sistemas típicos pueden ser muy elaborados. Un caso es el sistema de portadora tipo L4 de la Bell Telephone, en el cual se multicanalizan juntos 3 600 canales de voz (cada uno con $W = 4$ kHz en forma nominal) para transmisión por cable coaxial. Toda la modulación es en banda lateral única, tanto USSB como LSSB, y el espectro de banda base final va de 0.5 a 17.5 MHz, incluyendo la portadora piloto y las bandas de seguridad. Todas las subportadoras son múltiplos de 4 kHz y se derivan de un oscilador común, y la frecuencia piloto de 512 kHz proporciona la sincronización. Para evitar un excesivo ancho de la banda de seguridad en el extremos superior del espectro de banda base, se hace la multicanalización por *grupos* en cuatro etapas, un arreglo que también facilita la conmutación y el "enrutamiento" de los diferentes canales. (El lector que quizá no esté por completo convencido de la fuerza del análisis en el dominio de la frecuencia, podría considerar el diseño de ese sistema sin la ayuda de los conceptos espectrales.) Para referencia, la tabla 5.2 lista la jerarquía en FDM del Sistema Bell.‡

Ejemplo 5.4 La multicanalización en FM estéreo

Para concluir este capítulo, se describirá en forma breve el sistema FDM empleado en la radiodifusión de FM estereofónica comercial. Con referencia a la figura 5.37*a*, las señales de los altoparlantes izquierdo y derecho se *matrizan* primero para producir $x_l(t) + x_R(t)$ y $x_L(t) - x_R(t)$. (La señal suma se escucha con un receptor monofónico; se requiere del matrizado para que el oyente monaural no esté sometido a huecos de sonido en el material del programa y no capte así efectos estereofónicos de

† Así, la realimentación negativa es, por lo general, una necesidad en los sistemas de multicanalización por división en frecuencia para reducir la no linealidad de los amplificadores.

‡ Laboratorios Bell Telephone (1971, Cap. 6).

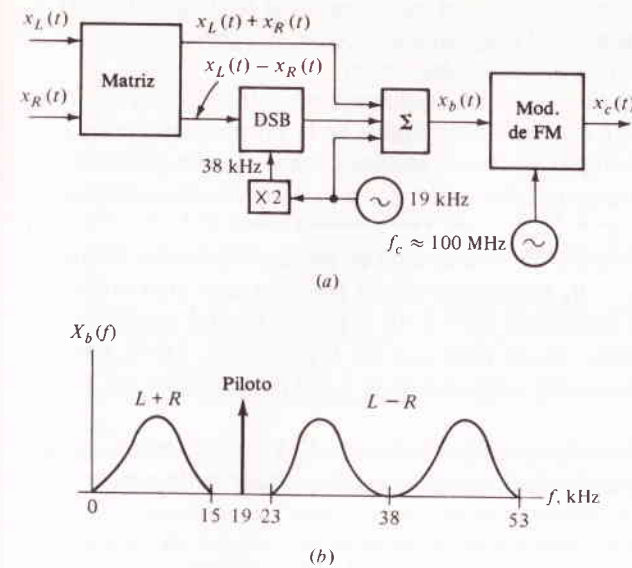



FIGURA 5.37
Multicanalización de FM estéreo. (*a*) Transmisor; (*b*) espectro de banda base.

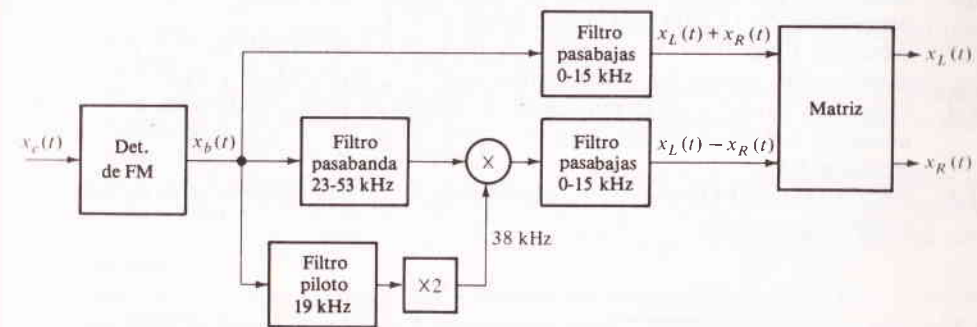



FIGURA 5.38
Receptor de FM estéreo multicanal.

ping pong.) La señal $x_L(t) + x_R(t)$ se inserta luego directamente en la banda base, en tanto que la DSB $x_L(t) - x_R(t)$ modula una subportadora de 38 kHz que se deriva de una fuente de 19 kHz. La modulación de doble banda lateral se emplea para mantener la fidelidad a las bajas frecuencias. El tono piloto de 19 kHz se agrega pa-

ra la sincronización del receptor, que aparece en el espectro de banda base (Fig. 5.37*b*). En la figura 5.38 se muestra el diagrama a bloques del receptor y cómo se emplea el tono piloto para la demodulación sincrónica.

De manera incidental, en la grabación de discos de cuatro canales discretos (sistema cuadrafónico) se emplea una extensión lógica de la estrategia anterior para multicanalizar cuatro señales independientes en los dos canales de una grabación estereofónica. Denotando las cuatro señales como L_F, L_R, R_F y R_R (para izquierda-frente, izquierda-atrás, etc.), $L_F + L_R$ se registra en forma directa en un canal con $L_F - L_R$ multicanalizadas por medio de modulación en frecuencia de una subportadora de 30 kHz. $R_F + R_R$ y $R_F - R_R$ se multicanalizan también en el otro canal. Como el espectro de banda base resultante sube a 45 kHz, las señales cuadrafónicas discretas no se pueden transmitir en su totalidad en FM estéreo. Otros sistemas cuadrafónicos† tienen sólo dos canales independientes y son compatibles con FM estéreo.

Hay aún mucho que hablar de la radiodifusión de FM estereofónica; se hará después de examinar las propiedades de la modulación de portadora, la cual es *exponencial*.

5.7 PROBLEMAS

5.1 (Secc. 5.1) Empleando un diagrama fasorial, encuéntrense expresiones para $v_i(t)$, $v_q(t)$, $R(t)$ y $\phi(t)$ cuando $v(t) = v_1(t)\cos[\omega_c t + \theta_1(t)] + v_2(t)\cos[\omega_c t + \theta_2(t)]$.
Respuesta: $v_i(t) = v_1(t)\cos\theta_1(t) + v_2(t)\cos\theta_2(t)$, $v_q(t) = v_1(t)\operatorname{sen}\theta_1(t) + v_2(t)\operatorname{sen}\theta_2(t)$.

5.2 (Secc. 5.1) Repítase el problema 5.1 para $v(t) = \cos\omega_c t + \alpha\cos(\omega_c + \omega_0)t$. Aproxímese $R(t)$ y $\phi(t)$ para el caso donde $\alpha^2 \ll 1$.

5.3 (Secc. 5.1) Cuando $z(t)$ es una señal pasabajas de banda limitada en $W < f_0$ emplee la ecuación (13), sección 2.3, y demuéstrese que $\int_{-\infty}^{\infty} z(t)\cos(\omega_0 t + \theta)\,dt = \theta$.

5.4 (Secc. 5.1) Si $v_i(t)$ y $v_q(t)$ son señales de energía de banda limitada en $W < f_c$, demuéstrese a partir de la ecuación (3) que $E_v = (E_{v_i} + V_{v_q})/2$. *Sugerencia*: Véase el problema 5.3.)

5.5 (Secc. 5.1) Dada $V(f)$ (Fig. P5.1), encuéntrense $v_{LP}(t)$ y la expresión de la portadora en cuadratura para $v(t)$.

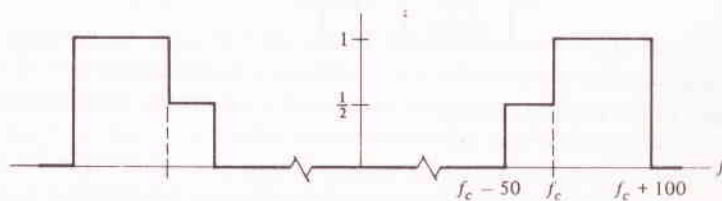



FIGURA P5-1

† Jurgen (1972) describe las diferentes aproximaciones a la cuadrafonía.

5.6 (Secc. 5.1) Cuando $Q \gg 1$, la función de transferencia del circuito sintonizado del ejemplo 2.3, sección 2.2, se puede aproximar a $H(f) \approx R/[1 + j(2Q/f_c)(f - f_c)$ para $f > 0$, donde $f_c = f_r$.
(*a*) Demuéstrese que $H_{LP}(f)$ tiene la forma de filtro pasabajas RC y encuéntrese $h(t)$ de $h_{LP}(t)$.
(*b*) Emplee la ecuación (9*b*) para encontrar la respuesta de pasabanda $y(t)$ cuando $x(t) = A\cos\omega_c t u(t)$.

5.7 (Secc. 5.1) El canal con $H_c(f)$ de la figura P4.1 se va a emplear para transmisión de pulsos en pasabanda, sin igualación. Selecciónese un valor apropiado para f_c y encuéntrese τ_{min}.

5.8 (Secc. 5.1) Se va a diseñar un sistema de radar para pulsos pasabanda de $1 - \mu$ s. ¿Cuál es el valor práctico más bajo para f_c?

5.9 (Secc. 5.2) Si $x(t) = 0.5\,u(t) - 1.5\,u(t - \tau)$, $\tau \gg 1/f_c$, bosquéjese $x_c(t)$ e indíquese la envolvente cuando la modulación es: AM con $m < 1$, AM con $m > 1$, de doble banda lateral.

5.10 (Secc. 5.2) La entrada a un sistema de AM con $m = 1$, es la señal moduladora de tono múltiple $x(t) = K(2\cos\omega_m t + \cos 2\omega_m t + 3\cos 5\omega_m t)$.
(*a*) Encuéntrese K para que $x(t)$ sea normalizada de manera apropiada, y constrúyase gráfica del espectro de líneas de frecuencia positiva de $x_c(t)$.
(*b*) Calcúlese P_{SB}/P_c y P_c/S_T.

5.11 (Secc. 5.2) Constrúyase el diagrama fasorial y encuéntrense $v_i(t)$, $v_q(t)$, y $R(t)$ para AM con modulación de tono a carga máxima cuando:
(*a*) Se atenúa la banda lateral superior en un factor de ½.
(*b*) Se le da a la banda lateral superior un corrimiento de fase de 180°.

5.12 (Secc. 5.2) Dedúzcase la ecuación (9) tomando el promedio del tiempo de $x_c^2(t)$. (*Sugerencia*: Véase el problema 5.3.)

5.13 (Secc. 5.2) Supóngase que $x(t)$ es una señal aleatoria gaussiana de valor medio cero donde la ecuación (2) de normalización no se puede aplicar. Tomando $m = 1$, encuéntrese el valor máximo de $\overline{x^2}$ tal que no haya sobremodulación durante más del 10% del tiempo. Calcúlese $2P_{SB}/S_T$ bajo esta condición.

5.14 (Secc. 5.3) Diséñese en forma de diagrama a bloques un modulador de AM, empleando un dispositivo no lineal que tenga $v_{salida} = av_{entrada}^5$. (*Sugerencia*: Se necesitará un multiplicador o un divisor de frecuencia.)

5.15 (Secc. 5.3) Supóngase que los moduladores de la figura 5.18 están *desbalanceados*; es decir, un dispositivo no lineal tiene $v_{salida} = a_1 v_{entrada} + a_2 v_{entrada}^2 + a_3 v_{entrada}^3$ mientras que el otro tiene $v_{salida} = b_1 v_{entrada} + b_2 v_{entrada}^2 + b_3 v_{entrada}^3$. Encuéntrese la señal de salida final.

5.16 (Secc. 5.3) Los moduladores tienen a menudo una *limitación de potencia pico*, así como un límite de potencia promedio. Encuéntrese $x_c^2(t)_{máx}$ en términos de $2P_{SB}$ para AM y DSB y compárese.

5.17 (Secc. 5.4) Empleando modulación de tono, demuéstrese que la salida en la figura 5.24 tiene la forma de la ecuación (2).

5.18 (Secc. 5.4) Sea $x(t)$ una señal de voz típica donde $X(f)$ 0 para $|f| < 200$ y $|f| > 3\,200$ Hz. Después de bosquejar los espectros en los puntos apropiados, demuéstrese que el sistema de la figura 5.2 genera modulación de banda lateral superior, y encuéntrense los máximos valores permitidos de f_{c_1} y f_{c_2} si las regiones de transición de los filtros pasaaltas deben satisfacer que $2\beta \geq 0.1\,f_{c0}$.

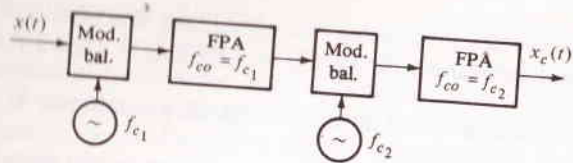



FIGURA P5-2

(Secc. 5.4) En la figura P5.3 se muestra un *modulador de banda lateral única de Weaver*. Analícese su operación empleando $x(t) = \cos 2\pi f_m t$, $f_m < W$.

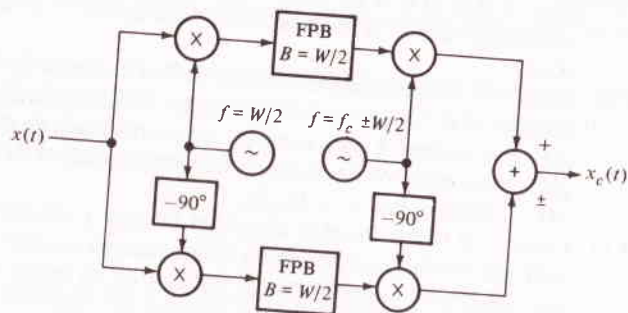



FIGURA P5-3

5.20★ (Secc. 5.4) Si $w(t) = \hat{v}(t)$, demuéstrese que $\hat{w}(t) = -v(t)$ y por medio de ello dedúzcase la transformada inversa de Hilbert $v(t) = -[\hat{v}(t)] * [1/\pi t]$.

5.21★ (Secc. 5.4) Compruébese la ecuación (13*b*).

5.22★ (Secc. 5.4) Cuando $x(t) = A\Pi(t/\tau)$, su transformada de Hilbert se puede escribir como

$$\hat{x}(t) = \frac{A}{\pi} \ln \left| \frac{t + \tau/2}{t - \tau/2} \right|$$

(*a*) Emplee la interpretación gráfica de la convolución para bosquejar a $\hat{x}(t) = [x(t)] * [l/\pi t]$ y demuéstrese que el bosquejo está de acuerdo con la expresión (163).
(*b*) Bosquéjese la envolvente de banda lateral única para este caso.

23 (Secc. 5.4) Obténgase una expresión para banda lateral residual con modulación de tono empleando $H(f_c + f_m) = 0.5 + a$ y $H(f_c - f_m) = 0.5 - a$. Escríbase la respuesta tanto en la forma de portadora en cuadratura como en la de envolvente y fase, y demuésrese que $x_c(t)$ se reduce a doble banda lateral o banda lateral única cuando $a = 0$ o ± 0.5, respectivamente.

4★ (Secc. 5.4) Demuéstrese que cuando se aplica una señal de doble banda lateral al filtro de banda lateral residual de la ecuación (14), resulta la ecuación (16). (*Sugerencia*: Empléense los equivalentes pasabajas para $H(f)$ y $x_c(t)$.)

5 (Secc. 5.4) Con modulación de tono, la modulación de banda lateral única más portadora (SSB + C) se puede escribir como $x_c(t) = A_c[\cos \omega_c t + a \cos(\omega_c \pm \omega_m)t]$.
(*a*) Elabórese el diagrama fasorial correspondiente y encuéntrese $R(t)$.

(*b*) Suponiendo $a \ll 1$, encuéntrese la condición en P_{SB}/P_c tal que la distorsión de envolvente por segunda armónica es menor del 5%.

5.26 (Secc. 5.4) Dedúzcase la ecuación (20) agregando una fase aleatoria Θ a ambos términos de portadora en la ecuación (18) y encontrando $E[x_c^2(t)]$.

5.27 (Secc. 5.5) El sistema de la figura P5.4 es un *perturbador de voz* simplificado que se emplea para asegurar la privacía en la comunicación y evitar la intercepción. Analícese su operación bosquejando el espectro en cada etapa, empleando $X(f)$ como se muestra. Demuéstrese también que bastará una unidad idéntica para funcionar como un desperturbador.

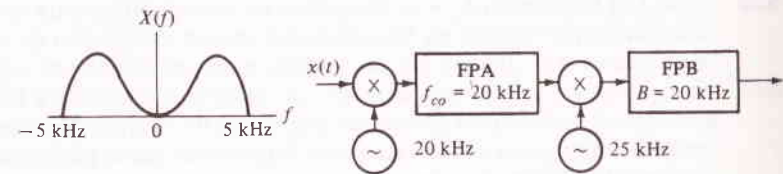



FIGURA P5-4

5.28 (Secc. 5.5) La figura P5.5 muestra un *divisor de frecuencia regenerativo*. Suponiendo una entrada senoidal de frecuencia f_0, demuéstrese que la frecuencia de salida es f_0/M, donde M es un entero. ¿Cuáles son las condiciones en el filtro pasabanda para esto?

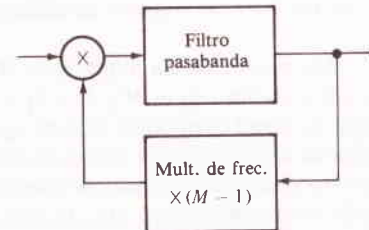



FIGURA P5-5

5.29 (Secc. 5.5) Empleando $x_c(t)$ como en la ecuación (2), encuéntrese $y_D(t)$ para detección sincrónica cuando la salida del oscilador local es $(\omega_c t + \phi)$, donde ϕ es un error de fase constante. Luego escríbanse respuestas separadas para modulación de amplitud, de doble banda lateral, de banda lateral única y de banda lateral residual por medio de una sustitución apropiada de k_c, k_m y $\zeta(t)$.

5.30 (Secc. 5.5) Supóngase que el diodo de la figura 5.30 tiene una *característica de ley cuadrática*. Encuéntrese y analícese la señal demodulada resultante.

5.31★ (Secc. 5.5) Con referencia al detector de envolvente de la figura 5.30, determínense los límites superior e inferior apropiados en función de la constante de tiempo R_1C_1 como queda determinado por la frecuencia portadora f_c y por la frecuencia moduladora máxima W. A partir de estos límites encuéntrese el valor práctico mínimo de f_c/W. (*Sugerencia*: Primero bosquéjese la onda rectificada para el caso de modulación de tono con $mA_m = 1$ y $f_m = W$.)

5.32 (Secc. 5.5) ¿Es posible diseñar una radio superheterodino de AM tal que la frecuencia imagen caiga siempre fuera de la banda de radiodifusión de AM? De serlo, ¿cuál es el valor necesario de f_{FI} y el intervalo de f_{OL}?

5.33 (Secc. 5.5) Dibújese y rotúlese en forma còmpleta el diagrama a bloques de un receptor superheterodino con las siguientes especificaciones: modulación de doble banda lateral con portadora piloto, $W = 15$ kHz, $f_c = 8$ a 10 MHz, $f_c' \leq 4$ MHz, detección homodina.

5.34 (Secc. 5.6) Se van a multicanalizar diez señales de voz, cada una de banda limitada a 3 kHz, y con 1 kHz de banda de seguridad entre canales. La modulación de subportadora es en banda lateral única y con $f_{c_1} = 0$.
(*a*) Calcúlese el ancho de banda de la banda base.
(*b*) Bosquéjese el espectro de la señal transmitida si la modulación de portadora es en AM.

5.35 (Secc. 5.6) Empleando $f_{c_1} = 0$, bosquéjese un espectro típico para un sistema de multicanalización por división en frecuencia que emplea modulación de subportadora y de portadora en AM. ¿Cuáles son las ventajas y desventajas de este esquema?

5.36 (Secc. 5.6) Con referencia a la tabla 5.2, dibújese el diagrama a bloques del sistema FDM que se necesita para generar un supergrupo de 5 grupos. Empléese modulación de banda lateral superior en todo el sistema. Suponiendo que se puede usar un banco de osciladores de subportadora para doce grupos, cuéntese el número total de osciladores y moduladores y compárese con el número que se requiere cuando todos los canales de voz se multicanalizan por división en frecuencia en un paso.

5.37 (Secc. 5.6) Los filtros pasabanda en un determinado receptor FDM tiene $H(f) = 10^{-(f - f_0)^2/2B^2}$ para $f > 0$, donde f_0 es la frecuencia central y B es de manera esencial el ancho de banda a 3 dB. Encuentre el ancho de banda de seguridad necesario en términos de B si se especifica que $|H(f)| \leq 10^{-1/2}$ en las regiones de rechazo para reducir al mínimo la diafonía.

5.38 (Secc. 5.6) Dos señales $x_1(t)$ y $x_2(t)$, cada una de banda limitada en W, se multicanalizan por división en frecuencia en las subportadoras $f_1 \gg W$ y $f_2 = f_1 + 2W$, respectivamente, y se transmiten sobre un canal no lineal cuya salida incluye $a_2 x_b^2(t) + a_3 x_b^3(t)$.
(*a*) Si la modulación de subportadora es de doble banda lateral, demuéstrese que toda la *diafonía* en la salida es *no inteligible* y que desaparece después del filtraje si $a_3 = 0$.
(*b*) Repítase para la modulación de subportadora en AM, la cual tendrá también diafonía *inteligible* debido a los términos de la forma $x_1(t) \cos \omega_2 t$.

6

MODULACIÓN EXPONENCIAL

Al empezar este capítulo se repiten dos propiedades de la modulación *lineal*: el espectro modulado es en forma básica el espectro del mensaje trasladado, y el ancho de banda de transmisión nunca excede en dos veces del mensaje. Una tercer propiedad, deducida en el capítulo 7, es que la relación señal a ruido en el destino no es mejor que en la transmisión en banda base y se puede mejorar sólo incrementando la potencia transmitida. La modulación exponencial difiere totalmente en estos tres enunciados.

En contraste con la modulación lineal, la modulación exponencial es un proceso *no lineal*; por lo que, no debe sorprender que el espectro modulado no esté relacionado con el espectro del mensaje de una manera sencilla. Más aún, resulta que el ancho de banda de transmisión es casi siempre mucho mayor que el doble del ancho de banda del mensaje. En compensación por la restricción del ancho de banda, existe el hecho que la modulación exponencial puede permitir el incremento de las relaciones señal a ruido sin que se tenga que incrementar la potencia transmitida.

Para obtener una apreciación cualitativa de esta propiedad, considérese una forma de modulación exponencial en donde la frecuencia instantánea de la onda modulada se varía de acuerdo con la forma de onda del mensaje, es decir, modulación en frecuencia. La señal demodulada es, por lo tanto, proporcional al intervalo de variación de frecuencia, conocido como desviación, y se puede incrementar la potencia de la señal de salida incrementando la desviación. Puesto que en este proceso

sólo se modula la frecuencia, la amplitud de la portadora permanece constante, y se produce una mejoría en la salida sin necesidad de incrementar la potencia de transmisión. Sin embargo, una desviación de frecuencia mayor requiere un ancho de banda de transmisión mayor. Así, con la modulación exponencial, se puede negociar ancho de banda por relación señal a ruido, una posibilidad embarazosa.

En forma irónica, la modulación de frecuencia se concibió primero como un medio de reducir el ancho de banda, siendo el argumento más o menos como sigue: si en vez de modular la amplitud de la portadora, se modula la frecuencia haciéndola oscilar en un intervalo de digamos ± 50 Hz, el ancho de banda de transmisión será entonces de 100 Hz, sin tomar en cuenta el ancho de banda del mensaje. Como pronto se verá, este argumento tiene un serio defecto; de manera específica, no considera la diferencia entre *frecuencia instantánea* y *frecuencia espectral*. Carlson (1922) reconoció la falacia de la noción de la reducción del ancho de banda y aclaró las cosas. De manera desafortunada, él y muchos otros también lamentaron que la modulación exponencial no presentara ventajas sobre la modulación lineal con respecto al ruido. Tomó algún tiempo superar esta creencia, pero gracias a Armstrong (1936), se apreciaron por fin los méritos de la modulación exponencial. Para poder entenderlos por completo, se deben considerar en forma cuidadosa los conceptos fundamentales.

6.1 CONCEPTOS FUNDAMENTALES

En la modulación exponencial, la onda modulada en forma fasorial es una función exponencial del mensaje, o sea,

$$x_c(t) = \text{Re}\,[A_c e^{j\theta_c(t)}] = A_c \cos \theta_c(t)$$

donde $\theta_c(t)$ es una función lineal de $x(t)$ y A_c es una constante. Puesto que θ_c es la posición angular del fasor, el nombre de modulación *angular* resulta igualmente apropiado.

Mientras que existen muchas formas posibles de modulación exponencial, sólo dos han demostrado ser prácticas, a saber, la *modulación de frecuencia* (FM) y la *modulación de fase* (PM). Estas designaciones sugieren frecuencia y fase variables en el tiempo, conceptos que requieren una interpretación especial. Esto es cierto en forma particular tratándose de la frecuencia variando en el tiempo, ya que la frecuencia implica periodicidad y la periodicidad variando en el tiempo es algo sin sentido.

Para aclarar el tema, se empieza por expresar a θ_c como

$$\theta_c(t) \triangleq 2\pi f_c t + \phi(t) \tag{1}$$

de manera que la frecuencia portadora f_c se especifica en forma unívoca. El segundo término de la ecuación (1) se puede interpretar como *ángulo de fase relativo,* en el sentido de que el fasor $e^{j\theta_c}$ difiere en posición angular de $e^{j\omega_c t}$ en $\phi(t)$. Recalcando más estas nociones, se debe recordar que la frecuencia angular (en radianes por segundo) es la derivada de la posición angular respecto al tiempo. Por lo que, se llega a

la definición de *desviación de frecuencia instantanea* $f(t)$ (en revoluciones o ciclos por segundo) por

$$f(t) \triangleq \frac{1}{2\pi}\frac{d\phi(t)}{dt} \tag{2}$$

lo cual puede se puede interpretar como la velocidad de $e^{j\theta_c}$ en comparación con $e^{j\omega_c t}$. Así, $\theta_c(t)$ está relacionada con $f(t)$ por integración en la forma

$$\theta_c(t) = 2\pi f_c t + 2\pi \int_{-\infty}^{t} f(\lambda)\, d\lambda \tag{3}$$

El límite inferior de integración representa un término de fase constante del que se puede prescindir sin pérdida de la generalidad si se desea. Con estos preliminares en mente, se define a la modulación de fase como el proceso en el que la fase es proporcional al mensaje, mientras que en la modulación en frecuencia quien es proporcional al mensajes, es la desviación de frecuencia instantánea $f(t)$.

De manera específica, la fase relativa de una onda de modulación de fase es

$$\phi(t) \triangleq \phi_\Delta x(t) \tag{4}$$

donde ϕ_Δ es la constante de desviación de fase, es decir, el máximo corrimiento de fase producido por $x(t)$ puesto que aún se está empleando la convención de mensaje $|x(t)| \leq 1$. La onda modulada es entonces

$$x_c(t) = A_c \cos\,[\omega_c t + \phi_\Delta x(t)] \tag{5}$$

En forma similar, la desviación de frecuencia instantánea de una onda de modulación de frecuencia es

$$f(t) \triangleq f_\Delta x(t) \tag{6}$$

donde f_Δ es la constante de desviación de frecuencia. Sustituyendo la ecuación (6) en la ecuación (3) se tiene

$$\theta_c(t) = 2\pi f_c t + 2\pi f_\Delta \int_{-\infty}^{t} x(\lambda)\, d\lambda \tag{7a}$$

y en consecuencia

$$x_c(t) = A_c \cos\left[\omega_c t + 2\pi f_\Delta \int_{-\infty}^{t} x(\lambda)\, d\lambda\right] \tag{7b}$$

es la forma de onda modulada

En lo anterior se supone que el mensaje no tiene componente de CD; o sea $\overline{x} = 0$. De otra manera la integral de la ecuación (7) divergiría conforme $t \to \infty$. En forma física, un término de CD en $x(t)$ produce un corrimiento de fase constante $\phi_\Delta \overline{x}$ en modulación de fase o un corrimiento de frecuencia portadora $f_\Delta \overline{x}$ en modulación de frecuencia. Prácticamente, cualquier componente de CD del mensaje por lo general se bloquea en los circuitos del modulador.

Tabla 6.1

	$\phi(t)$	$f(t)$
PM	$\phi_\Delta x(t)$	$\frac{\phi_\Delta}{2\pi}\frac{dx(t)}{dt}$
FM	$2\pi f_\Delta \int_{-\infty}^{t} x(\lambda)\,d\lambda$	$f_\Delta x(t)$

Comparando la ecuación (5) con la ecuación (7), se aprecia una leve diferencia entre la modulación de fase y la de frecuencia, siendo la diferencia esencial la integración del mensaje en FM. Más aún, no obstante la nomenclatura, ambas formas de modulación tienen fase y frecuencia variando en el tiempo, como lo subraya la Tabla 6.1. Estas relaciones indican en forma clara que, con la ayuda de redes de integración y diferenciación, un modulador de fase puede producir modulación de frecuencia y viceversa. En efecto, en el caso de la modulación de tono es virtualmente imposible distinguir en forma visual las ondas de modulación en frecuencia y modulación de fase.

Por otra parte, la comparación de la modulación exponencial con la modulación lineal revela algunas diferencias pronunciadas. Por un lado, la amplitud de una onda de modulación de frecuencia o de fase es siempre constante; por lo que, sin tomar en cuenta el mensaje $x(t)$, la potencia transmitida promedio es†

$$S_T = \tfrac{1}{2}A_c^{\,2} \tag{8}$$

Por otro, los cruzamientos cero de una onda modulada en forma exponencial no son periódicos, pero lo son siempre en la modulación lineal. Además, por la propiedad de amplitud constante de las modulaciones de frecuencia y de fase, se puede decir que el mensaje reside *sólo en los cruzamientos cero,* siempre y cuando la frecuencia portadora sea grande.‡ Por último, puesto que la modulación exponencial es un proceso no lineal, la onda modulada no se parece en nada a la forma de onda del mensaje.

La figura 6.1 ilustra algunos de estos puntos mostrando las ondas típicas de AM, FM y PM. Si el lector desea puede, como ejercicio, confrontar estas formas de onda con las señales moduladoras correspondientes. En FM y PM esto se hace con más facilidad considerando la frecuencia instantánea, mejor aún que sustituyendo a $x(t)$ en las ecuaciones (5) y (7).

No obstante las muchas similitudes entre la FM y la PM, la modulación de frecuencia resulta tener propiedades de reducción de ruido superiores por lo que se le concederá mayor atención. Por fortuna, los resultados y conclusiones que se obtienen de un estudio de la modulación de frecuencia son aplicables aunque con modificaciones menores a todas las formas de la modulación exponencial.

† Existen excepciones teóricas para esta condición cuando $f(t) \leq -f_c$, pero ellas ocurren raramente en la práctica.

‡ Schwartz demuestra esto (1970, Cap. 4).

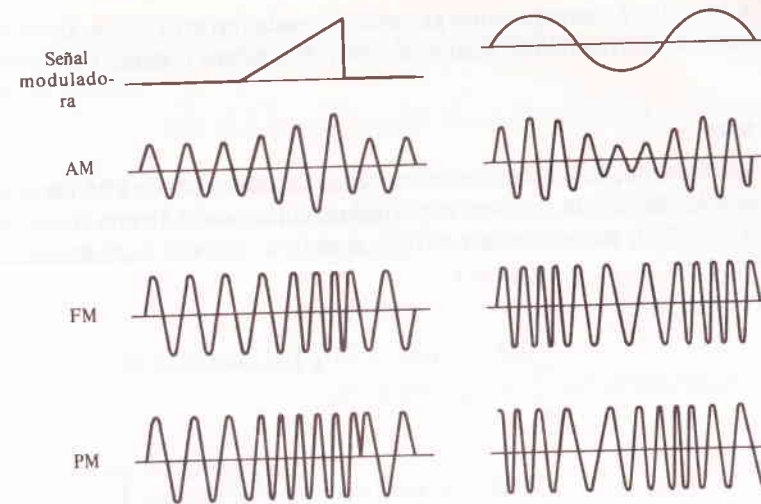



FIGURA 6.1
Formas de ondas ilustrativas de AM, FM y PM.

EJERCICIO 6.1 Supóngase que la FM se definiera por analogía directa con la AM como $x_c(t) = A_c \cos \omega_c[1 + mx(t)]t$. Demuéstrese la imposibilidad física de esta definición encontrando $f(t)$ cuando $x(t) = \cos \omega_m t$. *Resp.:* $mf_c[\cos \omega_m t - \omega_m t \operatorname{sen} \omega_m t]$.

6.2 ANÁLISIS ESPECTRAL DE FM

La ecuación (7), (Sec. 6.1) da la descripción del dominio del tiempo de una onda de FM con mensaje arbitrario $x(t)$. En consecuencia, se empieza este estudio con la descripción del dominio de la frecuencia, o análisis espectral, de FM. Antes de hacerlo, es necesario observar que $f(t)$ no es lo mismo que la *frecuencia espectral f*. La primera es una cantidad dependiente del tiempo que describe a $x_c(t)$ en el dominio del tiempo; la última es la variable independiente del análisis espectral en donde $X_c(f)$ describe a $x_c(t)$ en el dominio de la frecuencia en términos de las componentes senoidales de frecuencia fija. Por lo que no se puede esperar una correspondencia simple uno a uno entre $f(t)$ y el espectro de FM.

Como implican estas consideraciones, es difícil una descripción exacta de los espectros de FM salvo para ciertas señales moduladoras sencillas. (Esto, por supuesto, simplemente refleja el hecho de que la modulación exponencial es un proceso *no lineal.*) Por tanto, en vez de intentar el análisis con una $x(t)$ arbitraria, se examinarán varios casos específicos, como una táctica alterna, empezando con la modula-

ción de tono, y se formularán conclusiones generales basadas en esos casos. De antemano se acepta que este acercamiento ocurre después de algunos rodeos, pero no es necesario.

Modulación de tono

Con la modulación de tono, la frecuencia instantánea de una señal de FM varía en forma senoidal en relación con la frecuencia portadora (ilustración de una forma de onda típica, en la Fig. 6.1), de manera específica, si $x(t) = A_m \cos \omega_m t$, entonces, omitiendo el límite inferior de integración,

$$\theta_c(t) = 2\pi f_c t + 2\pi f_\Delta \int^t A_m \cos \omega_m \lambda \, d\lambda$$

y

$$x_c(t) = A_c \cos\left(\omega_c t + \frac{2\pi f_\Delta A_m}{\omega_m} \operatorname{sen} \omega_m t\right)$$

Para simplificar la notación, sea

$$\beta \triangleq \frac{2\pi f_\Delta A_m}{\omega_m} = \frac{A_m f_\Delta}{f_m} \tag{1}$$

tal que

$$x_c(t) = A_c \cos(\omega_c t + \beta \operatorname{sen} \omega_m t) \tag{2}$$

Este resultado debe contrastar con la ecuación (11), (Sec. 5.2), la expresión equivalente para AM con modulación de tono.

El parámetro β que se introdujo antes, se conoce como *índice de modulación* de FM, y presenta dos propiedades extraordinarias: está definido sólo para modulación de tono, y depende tanto de la amplitud como de la frecuencia del tono modulante. En forma física, β es la *desviación de fase máxima* (en radianes) producida por el tono en cuestión. Esta conclusión se desprende de la inspección de la ecuación (2), la cual muestra que la fase relativa de $x_c(t)$ es

$$\phi(t) = \beta \operatorname{sen} \omega_m t = \frac{A_m f_\Delta}{f_m} \operatorname{sen} \omega_m t \tag{3}$$

Así, diferentes tonos con la misma relación amplitud a frecuencia ocasionan la misma desviación de fase, pero a diferentes regímenes. Sin embargo, puesto que $f(t) = f_\Delta A_m \cos \omega_m t$, la desviación de frecuencia depende sólo de la amplitud del tono y f_Δ, viene a ser la última propiedad del modulador.

Para el análisis espectral de $x_c(t)$, no se intentará una transformación de Fourier directa de la ecuación (2), por razones más que obvias. Pero es posible expresar $x_c(t)$ como una suma de sinusoides, lo cual nos da entonces el espectro de

líneas de frecuencia positiva. Para este propósito se escribe primero la ecuación (2) en la forma

$$x_c(t) = A_c[\cos \omega_c t \cos(\beta \operatorname{sen} \omega_m t) - \operatorname{sen} \omega_c t \operatorname{sen}(\beta \operatorname{sen} \omega_m t)] \tag{4}$$

y obsérvese que, aun cuando $x_c(t)$ en sí misma no es necesariamente periódica $\cos(\beta \operatorname{sen} \omega_m t)$ y $\operatorname{sen}(\beta \operatorname{sen} \omega_m t)$ son periódicas en $1/f_m$, pudiéndose así desarrollar por medio de serie de Fourier. En particular, es bien conocido en matemáticas aplicadas que

$$\begin{aligned} \cos(\beta \operatorname{sen} \omega_m t) &= J_0(\beta) + \sum_{n \text{ par}}^{\infty} 2J_n(\beta) \cos n\omega_m t \\ \operatorname{sen}(\beta \operatorname{sen} \omega_m t) &= \sum_{n \text{ impar}}^{\infty} 2J_n(\beta) \operatorname{sen} n\omega_m t \end{aligned} \tag{5}$$

donde n es positiva y

$$J_n(\beta) \triangleq \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j(\beta \operatorname{sen} \lambda - n\lambda)} \, d\lambda \tag{6}$$

Los coeficientes $J_n(\beta)$ son *funciones de Bessel* de primera clase, de orden n y argumento β. Con la ayuda de la ecuación (6), el lector puede reducir las dificultades al efectuar los desarrollos trigonométricos de la ecuación (5).

Sustituyendo la ecuación (5) en la ecuación (4) y desarrollando los productos de senos y cosenos se tiene por último

$$\begin{aligned} x_c(t) = {} & A_c J_0(\beta) \cos \omega_c t \\ & + \sum_{n \text{ impar}}^{\infty} A_c J_n(\beta)[\cos(\omega_c + n\omega_m)t - \cos(\omega_c - n\omega_m)t] \\ & + \sum_{n \text{ par}}^{\infty} A_c J_n(\beta)[\cos(\omega_c + n\omega_m)t + \cos(\omega_c - n\omega_m)t] \end{aligned} \tag{7a}$$

En forma alterna, aprovechando la ventaja de la propiedad de que $J_{-n}(\beta) = (-1)^n J_n(\beta)$ se obtiene la más compacta pero menos informativa expresión

$$x_c(t) = A_c \sum_{n=-\infty}^{\infty} J_n(\beta) \cos(\omega_c + n\omega_m)t \tag{7b}$$

En una u otra forma, la ecuación (7) es la representación matemática de una onda de amplitud constante cuya frecuencia instantánea está variando en forma sinusoidal. En breve se dará una interpretación fasorial, que arrojará más luz sobre el tema.

Examinando la ecuación (7), se ve que al espectro de FM lo forman una línea correspondiente a la frecuencia portadora, más un número *infinito* de líneas corres-

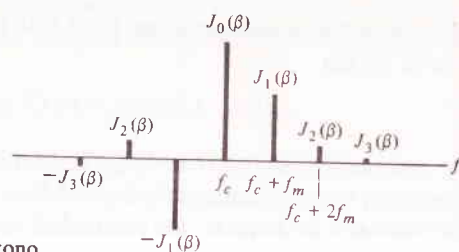



FIGURA 6.2
Espectro de líneas de FM, modulación de tono.

pondientes a las bandas laterales a frecuencias † $f_c \pm nf_m$. Como se ilustra en el espectro típico (Fig. 6.2), todas las líneas están igualmente espaciadas en un valor igual al de la frecuencia moduladora y las de banda lateral inferior de orden impar, están invertidas en fase (es decir, tienen amplitudes "negativas") en comparación con la portadora no modulada. En general, la amplitud relativa de una línea en $f_c + nf_m$ está dada por $J_n(\beta)$, pero antes de que se pueda decir algo más acerca del espectro, se debe examinar el comportamiento de las funciones de Bessel.

La figura 6.3 muestra unas cuantas funciones de Bessel de diferentes órdenes cuyas gráficas están en función del argumento β. En esta gráfica se pueden advertir varias propiedades importantes.

1. La amplitud relativa de la línea portadora $J_0(\beta)$ varía con el índice de modulación y, en consecuencia, depende de la señal moduladora. Así, en contraste con la moculación lineal, la componente de frecuencia portadora de una onda de FM "contiene" parte de la información del mensaje. No obstante,

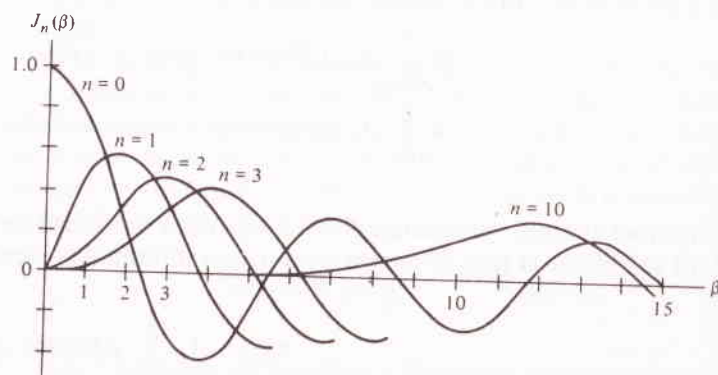



FIGURA 6.3
Funciones de Bessel de orden fijo graficadas contra el argumento β.

† Se está tratando con un espectro de líneas de frecuencia positiva, pues las frecuencias negativas aparentes debidas a $nf_m > f_c$ están plegadas hacia los valores positivos $|f_c - nf_m|$. Tales componentes generalmente se desprecian en la práctica cuando la frecuencia portadora es mayor en muchos órdenes que la frecuencia moduladora.

habrá espectros en los cuales la línea de portadora tenga amplitud cero, puesto que $J_0(\beta) = 0$ cuando $\beta = 2.4, 5.5$, etc.

2. El número de líneas de banda lateral que tienen amplitud relativa apreciable, también es una función de β. Con $\beta \ll 1$ sólo J_0 o J_1 son significativas, por lo que el espectro consistirá de líneas portadoras y de dos bandas laterales, muy parecido al de AM, salvo por la inversión de fase de línea de la banda lateral inferior. Por otra parte, si $\beta \gg 1$, habrá muchas líneas de banda lateral, dando al espectro bastante diferencia respecto de la modulación lineal.

3. Un índice de modulación β grande implica un ancho de banda grande para acomodar la vasta estructura de bandas laterales en concordancia con la interpretación física de gran desviación de frecuencia.

Algunos de los puntos anteriores se ilustran mejor en la figura 6.4, la cual da a $J_n(\beta)$ como una función de n/β para diferentes valores *fijos* de β. Puesto que en la FM con modulación de tono se tiene a β constante, estas curvas representan la "envolvente" de las líneas de banda lateral si se multiplica el eje horizontal por βf_m para obtener la posición de la línea nf_m relativa a f_c. Obsérvese en particular que todas las $J_n(\beta)$ decaen en forma monótona para $n/\beta > 1$ y que $|J_n(\beta)| \ll 1$ si $|n/\beta| \gg 1$. En forma similar a la figura 6.4, la tabla 6.2 lista valores seleccionados de $J_n(\beta)$, aproximados a los decimales†. Los espacios en blanco en la tabla corresponden a $|J_n(\beta)| < 0.01$.

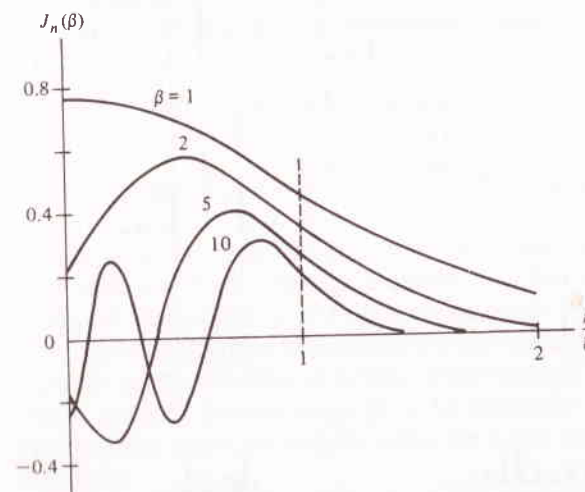



FIGURA 6.4
Funciones de Bessel de argumento fijo graficadas contra n/β.

† En Jahnke y Emde (1945), se dan tabulaciones más extensas.

Tabla 6.2 VALORES SELECCIONADOS DE $J_n(\beta)$

n	$J_n(0.1)$	$J_n(0.2)$	$J_n(0.5)$	$J_n(1.0)$	$J_n(2.0)$	$J_n(5.0)$	$J_n(10)$	n
0	1.00	0.99	0.94	0.77	0.22	−0.18	−0.25	0
1	0.05	0.10	0.24	0.44	0.58	−0.33	0.04	1
2			0.03	0.11	0.35	0.05	0.25	2
3				0.02	0.13	0.36	0.06	3
4					0.03	0.39	−0.22	4
5						0.26	−0.23	5
6						0.13	−0.01	6
7						0.05	0.22	7
8						0.02	0.32	8
9							0.29	9
10							0.21	10
11							0.12	11
12							0.06	12
13							0.03	13
14							0.01	14

En la figura 6.5 se muestran espectros de líneas típicos, en los cuales, para claridad, se han omitido las inversiones de las líneas de banda lateral inferior de orden impar. Estos espectros deben ser examinados en forma cuidadosa por el lector dada la influencia relativa de la amplitud y frecuencia modulantes. Nótese también la concentración del espectro dentro de $f_c \pm \beta f_m$ cuando β es grande.

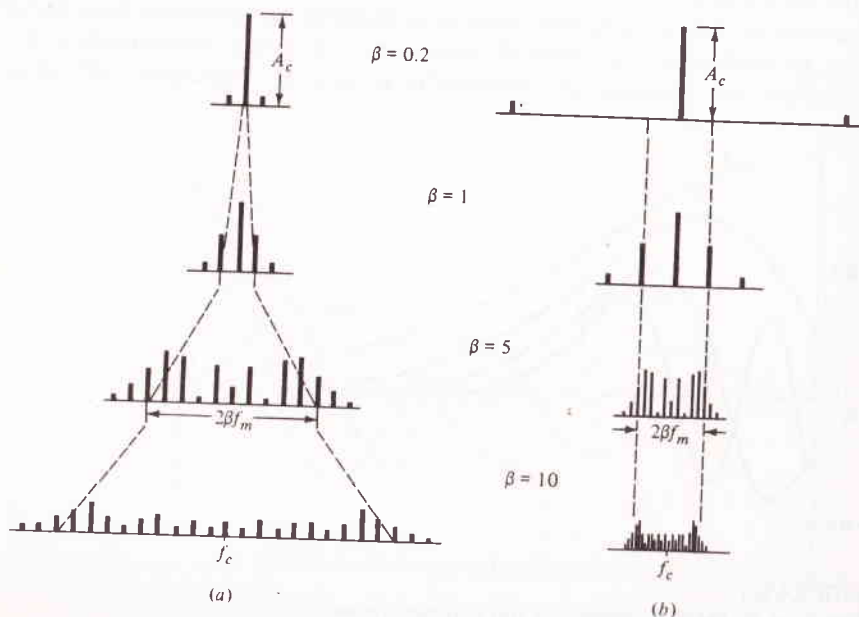



FIGURA 6.5
Espectros de líneas de FM de tono modulado. (a) f_m fija, $A_m f_\Delta$ creciente; (b) $A_m f_\Delta$ fija, f_m decreciente.

EJERCICIO 6.2 Dibújese el espectro de líneas cuando f_c = 12 kHz, f_m = 2 kHz y $A_m f_\Delta$ = 4 kHz. Si $A_m f_\Delta$ = 10 kHz, demuéstrese que $x_c(t)$ tiene una componente de CD y en consecuencia $S_T \neq A_c^2/2$.

Interpretación fasorial

Dado que la forma de expresar a $x_c(t)$ en la ecuación (7) es bastante engorrosa, se recurre al diagrama fasorial de la FM como ayuda en la interpretación física. Como punto de partida, supóngase que $\beta \ll 1$, de tal manera que $J_c(\beta) \approx 1, J_1(\beta) \approx \beta/2$, y se hace caso omiso de todas las líneas de orden superior. Esto hace que el diagrama fasorial de la figura 6.6, sea uno que difiere de aquel del caso de la AM (Fig. 5.9) sólo en la inversión de fase de la línea de banda lateral inferior. Pero a causa de la inversión de fase, la contribución del par de banda lateral es más bien perpendicular o *en cuadratura* respecto a la portadora,

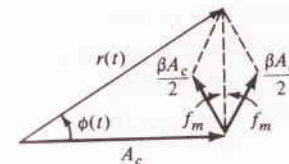



FIGURA 6.6
Diagrama fasorial de FM para $\beta \ll 1$.

que colineal. Esta relación de cuadratura es precisamente lo que se necesita para producir modulación de fase o de frecuencia en vez de modulación en amplitud.

En forma analítica, la envolvente y fase de $x_c(t)$ con β pequeña son

$$R(t) \approx \sqrt{A_c^2 + \left(2\frac{\beta}{2}A_c \operatorname{sen}\omega_m t\right)^2} \approx A_c\left[1 + \frac{\beta^2}{4} - \frac{\beta^2}{4}\cos 2\omega_m t\right] \tag{8}$$

$$\phi(t) \approx \arctan\left[\frac{2(\beta/2)A_c \operatorname{sen}\omega_m t}{A_c}\right] \approx \beta \operatorname{sen}\omega_m t$$

Así, la variación de fase es en forma aproximada como se desea, pero existe una variación de *amplitud* adicional al doble de la frecuencia de tono. Para cancelar esto último, se debe incluir el par de líneas de banda lateral de segundo orden que giran a $\pm 2f_m$ en relación con la portadora y cuya resultante es colineal con ella. Dado que el par de segundo orden en forma virtual concela la modulación de amplitud no deseada, también distorsiona a $\phi(t)$. La distorsión se corrige entonces agregando el par de tercer orden, lo cual introduce de nuevo modulación en amplitud, y así hasta el infinito.

Cuando se incluyen todas las líneas espectrales, los pares de orden non tienen una resultante en cuadratura con la portadora que provee la modulación en frecuencia deseada más una modulación en amplitud no deseada. La resultante de los pares de orden par, que son colineales con la portadora, corrigen las variaciones de amplitud. El efecto neto es entonces como el que se muestra en la figura 6.7. El extremo de los barridos resultantes en un arco circular, refleja la amplitud constante A_c.

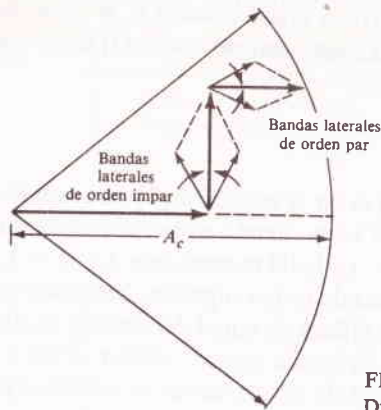



FIGURA 6.7
Diagrama fasorial de FM para β arbitrario.

Modulación de tono múltiple

La técnica de serie de Fourier que se usó para llegar a la ecuación (7) se puede aplicar también al caso de la modulación de tono múltiple. Por ejemplo, supóngase que $x(t) = A_1 \cos \omega_1 t + A_2 \cos \omega_2 t$, donde f_1 y f_2 no están relacionadas en forma armónica. La onda modulada se escribe primero como

$$x_c(t) = A_c[\cos \omega_c t(\cos \alpha_1 \cos \alpha_2 - \operatorname{sen}\alpha_1 \operatorname{sen}\alpha_2) - \operatorname{sen}\omega_c t(\operatorname{sen}\alpha_1 \cos \alpha_2 + \cos \alpha_1 \operatorname{sen}\alpha_2)]$$

donde $\alpha_1 = \beta_1 \operatorname{sen} \omega_1 t$, $B_1 = A_1 f_\Delta / f_1$, etc. Los términos de la forma $\cos \alpha_1$, $\operatorname{sen} \alpha_1$, etc., se desarrollan luego conforme a la ecuación (5), y después de algunas manipulaciones rutinarias, se llega al resultado simplificado.

$$x_c(t) = A_c \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} J_n(\beta_1) J_m(\beta_2) \cos(\omega_c + n\omega_1 + m\omega_2)t \tag{9}$$

Interpretando esta expresión en el dominio de la frecuencia, las líneas espectrales se pueden dividir en cuatro categorías: (1) la línea portadora de amplitud

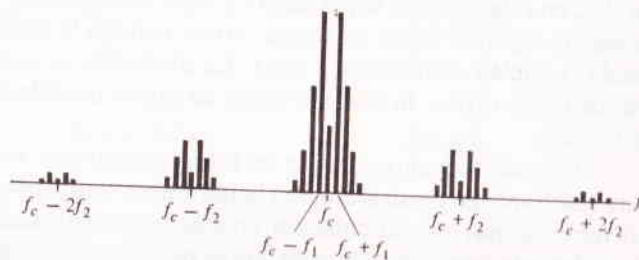



FIGURA 6.8
Espectro de línea de FM de doble tono, $f_1 \ll f_2$, $\beta_1 \geq \beta_2$.

$A_c J_0(\beta_1) J_0(\beta_2)$; (2) las líneas de banda lateral en $f_c \pm nf_1$ debidas a un tono solo; (3) las líneas de banda lateral en $f_c \pm mf_2$ debidas al otro tono solo; y (4) las líneas de banda lateral en $f_c \pm nf_1 \pm mf_2$ las cuales aparecen como modulación de frecuencia debatido en las frecuencias suma y diferencia de los tonos modulantes y sus armónicas. Esta última categoría puede caer de sorpresa, porque no viene paralela a la modulación lineal donde la norma es la simple superposición de las líneas de banda lateral. Pero entonces se debe recordar que la FM es una modulación *no lineal*, donde no se espera la superposición. En la figura 6.8, se muestra un espectro de FM de doble tono en el que se ven diferentes tipos de líneas espectrales para $f_1 \ll f_2$ y $\beta_1 > \beta_2$. En estas condiciones existe la curiosa propiedad de que cada línea de banda lateral en $f_c \pm mf_2$ se ve como otra portadora de FM con modulación de tono de frecuencia f_1.

Si es necesario en forma absoluta, la técnica anterior se puede hacer extensiva a la modulación con más de dos tonos no armónicos; el procedimiento es directo pero desordenado.

Cuando las frecuencias de tono se relacionan en forma armónica, es decir, que $x(t)$ es una forma de onda *periódica*, entonces $\phi(t)$ es periódica y, por lo tanto, es $e^{j\phi(t)}$. Lo último se puede desarrollar en serie exponencial de Fourier con coeficientes

$$c_n = \frac{1}{T_0} \int_{T_0} \exp j[\phi(t) - n\omega_0 t]\, dt \tag{10a}$$

Por lo que

$$x_c(t) = \operatorname{Re}\left[A_c \sum_{n=-\infty}^{\infty} c_n e^{j(\omega_c + n\omega_0)t}\right] \tag{10b}$$

Se omite aquí un examen más a fondo, puesto que ello no agrega nada que tenga una importancia particular†.

Modulación de pulsos

No obstante las complejidades de los espectros de FM, hay unas cuantas señales moduladoras para las cuales $x_c(t)$ se puede tratar por medio de transformaciones de Fourier directas. Una de tales señales es el pulso rectangular $x(t) = \Pi(t/\tau)$, que es un ejemplo importante e informativo.

La clave de la transformación es la frecuencia instantánea puesto que, para la modulación de pulsos, es en forma simple $f(t) = f_\Delta \Pi(t/\tau)$. Por inspección o integración de $f(t)$ se tiene

$$x_c(t) = \begin{cases} A_c \cos(\omega_c + \omega_\Delta)t & |t| < \dfrac{\tau}{2} \\ A_c \cos \omega_c t & |t| > \dfrac{\tau}{2} \end{cases}$$

† Panter (1965, Cap. 7) da detalles más amplios y varios ejemplos interesantes.

Pero es más apropiada para lo que se persigue, la forma

$$x_c(t) = A_c\left[\cos \omega_c t - \prod\left(\frac{t}{\tau}\right) \cos \omega_c t + \prod\left(\frac{t}{\tau}\right) \cos (\omega_c + \omega_\Delta)t\right] \quad (11)$$

la cual dice que la señal modulada es una onda sinusoidal de frecuencia f_c menos un pulso de RF de frecuencia f_c más un pulso de RF de frecuencia $f_c + f_\Delta$.

Aunque la ecuación (11) puede parecer indebidamente formal, se presta por sí misma con facilidad al análisis de Fourier, además, con el resultado del ejemplo 2.7 (Sec. 2.3), se puede transformar a $x(t)$ término por término para dar

$$X_c(f) = \frac{A_c}{2}[\delta(f-f_c) + \delta(f+f_c)] - \frac{A_c\tau}{2}[\text{senc}(f-f_c)\tau + \text{senc}(f+f_c)\tau]$$
$$+ \frac{A_c\tau}{2}[\text{senc}(f-f_c-f_\Delta)\tau + \text{senc}(f+f_c+f_\Delta)\tau] \quad (12)$$

La porción de frecuencia positiva de $|X_c(f)|$ se bosqueja en la figura 6.9 para $f_\Delta = 2/\tau$. Se ve que el espectro *no es simétrico* en relación con la frecuencia portadora y tiene más contenido encima de f_c que por debajo; un hecho que se podría haber anticipado porque $f_c + f(t)$ nunca es menor que f_c. Más aún, no obstante que están presentes otras frecuencias, el espectro se concentra en los dos valores discretos f_c y $f_c + f_\Delta$.

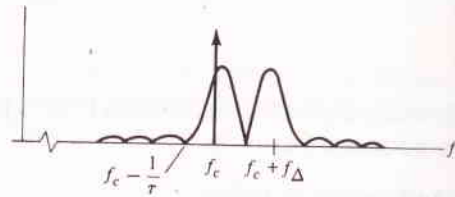



FIGURA 6.9
Espectro de FM de pulso modulado, $f_\Delta = 2/\tau$.

6.3 ANCHOS DE BANDA EN FM

Se ha visto que, en general, un espectro de FM tiene extensión infinita. En consecuencia, la generación y transmisión de FM *pura* necesita sistemas de *ancho de banda infinito*, sea el mensaje o no de banda limitada. Pero los sistemas prácticos de FM con ancho de banda finito sí existen y se desempeñan bastante bien. Su éxito depende de que, suficientemente lejos de la frecuencia portadora, las componentes espectrales son muy pequeñas y se pueden descartar. Cierto, omitiendo cualquier porción del espectro se dará lugar a la *distorsión* de la señal demodulada; pero la distorsión se puede reducir al mínimo conservando todas las componentes espectrales *significativas*.

La determinación del ancho de banda de transmisión en FM se reduce a la pregunta: ¿Qué tanto del espectro de la señal modulada es significativo? Por supuesto

que las normas de significación no son absolutas, siendo contingentes con la cantidad de distorsión que se puede tolerar en una aplicación específica. Sin embargo, con los criterios de las reglas de uso inmediato basadas en el estudio de la modulación de tono, se ha logrado un éxito considerable y se ha llegado a relaciones aproximadas útiles. El análisis de los requisitos de ancho de banda de FM empieza, por lo tanto, con las líneas de banda lateral significativas para modulación de tono.

Líneas de banda lateral significativas

La figura 6.4 indica que $J_n(\beta)$ decae en forma rápida para $|n/\beta| > 1$, particularmente si $\beta \gg 1$. Suponiendo que el índice de modulación β es grande, se puede decir que $|J_n(\beta)|$ es significativo sólo para $|n| \leq \beta = A_m f_\Delta / f_m$. Por lo que, todas las líneas de banda lateral significativas están contenidas en el intervalo de frecuencia $f_c \pm \beta f_m = f_c \pm A_m f\Delta$; una conclusión en concordancia con el razonamiento intuitivo. Por otra parte, supóngase que el índice de modulación es pequeño; entonces *todas* las líneas de banda lateral son pequeñas comparadas con la portadora, puesto que $J_0(\beta) \gg J_{n \neq 0}(\beta)$ cuando $\beta \ll 1$. Sin embargo, se deben conservar, por lo menos, el par de bandas laterles de primer orden, ya que de otra manera no habría modulación en frecuencia. En consecuencia para β pequeño, las líneas de banda lateral significativas están contenidas en $f_c \pm f_m$.

Para colocar las observaciones anteriores sobre una base cuantitativa, todas las líneas de banda lateral que tienen como amplitud relativa $|J_n(\beta)| > \epsilon$ están *definidas* como significativas, donde ϵ va de 0.01 a 0.1 de acuerdo con la aplicación. En-

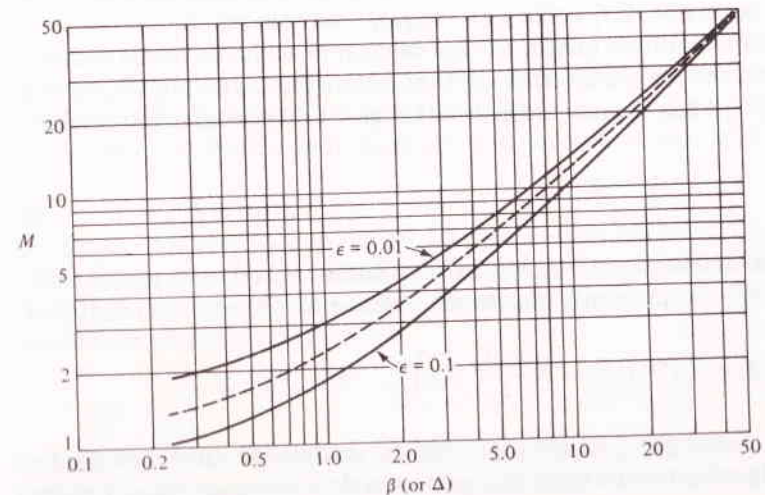



FIGURA 6.10
El número de pares de bandas laterales significativas como una función de β (o Δ).

tonces, si $|J_M(\beta)| > \epsilon$ y $|J_{M+1}(\beta)| < \epsilon$, existen M *pares* de bandas laterales y $2M + 1$ líneas significativas. Así, el ancho de banda es

$$B = 2M(\beta)f_m \qquad M \geq 1 \tag{1}$$

puesto que las líneas están espaciadas en f_m y M depende del índice de modulación β. En la ecuación (1) se ha incluido la condición $M \geq 1$ para dar fe de que B no puede ser menor que $2f_m$.

La figura 6.10 muestra a M como una función continua de β para $\epsilon = 0.01$ y 0.1. Estudios experimentales indican que lo primero es a menudo bastante conservativo, mientras que lo último puede dar por resultado distorsión pequeña pero perceptible. Son aceptables para muchos propósitos, los valores de M entre estos dos límites —conforme se indica por medio de la línea punteada— y se emplearán en lo sucesivo.

Pero el ancho de banda B no es el ancho de banda de transmisión B_T; más bien es el ancho de banda mínimo necesario para modulación por medio de un tono de amplitud y frecuencia específicas. Como ilustración, la máxima desviación de frecuencia f_Δ en la FM comercial está limitada por la FCC a 75 kHz, y las frecuencias moduladoras cubren de 30 Hz a 15 kHz. Si un tono de 15 kHz tiene amplitud unitaria ($A_m = 1$), entonces $\beta = 75/15 = 5$, $M = 7$, y $B = 2 \times 7 \times 15 = 210$ kHz. Si la amplitud hubiera sido menor, no se hubiera desarrollado la máxima desviación de frecuencia, y en ancho de banda sería más pequeño†. Más aún, un tono de baja frecuencia, digamos de 7.5 kHz a amplitud plena, resultaría en un mayor índice de modulación ($\beta = 10$), o sea, un número más grande de pares de bandas laterales significatifas ($M = 12$), pero un ancho de banda más pequeño, a saber, $B = 2 \times 12 \times 7.5 = 180$ KHz. En pocas palabras, el ancho de banda se determina en una forma más que compleja tanto por $A_m f_\Delta$ como f_m(o β y f_m), y no sólo por β.

Para conseguir este último punto, se debe calcular el ancho de banda *máximo* que se requiere cuando los parámetros de tono están limitados por $A_m \leq 1$ y $f_m \leq W$. Para este propósito, la línea punteada de la figura 6.10 se puede aproximar por

$$M(\beta) \approx \beta + \alpha \tag{2}$$

donde α es en forma esencial constante con un valor entre 1 y 2 (el valor exacto carece de importancia por el momento). Insertando la ecuación (2) en la (1) se tiene

$$B \approx 2(\beta + \alpha)f_m = 2\left(\frac{A_m f_\Delta}{f_m} + \alpha\right)f_m = 2(A_m f_\Delta + \alpha f_m)$$

Ahora, teniendo en cuenta que f_Δ es una propiedad del modulador, ¿qué tono produce el ancho de banda máximo? Es claro que es el tono *de máxima amplitud* y *máxima frecuencia* con $A_m = 1$ y $f_m = W$. Entonces, el ancho de banda en modulación de tono, en el peor de los casos es

$$B_{\text{máx}} = 2(f_\Delta + \alpha W) \tag{3}$$

Nótese en forma cuidadosa que el correspondiente índice de modulación $\beta = f_\Delta / W$ no es el valor máximo de β sino más bien el valor que, combinado con la máxima frecuencia moduladora, da lugar al ancho de banda máximo. Cualquier otro tono que tenga $A_m < 1$ o $f_m < W$ requerirá menos ancho de banda aunque β sea más grande.

† Resulta paradójico que en modulación *de frecuencia* (con f_Δ fija) el ancho de banda dependa de la *amplitud* el ancho de banda dependa de la *frecuencia* del tono.

Ancho de banda de transmisión

Se enfoca ahora la atención en el ancho de banda de transmisión B_T que se necesita cuando $x(t)$ es una *señal moduladora arbitraria* con el ancho de banda W y que satisface la convención de normalización $|x(t)| \leq 1$. Pero no hay que volver la espalda a las conclusiones previas; en efecto, se estimará de B_T en forma directa a partir del análisis de la modulación de tono en el peor de los casos, suponiendo que cualquier componente de $x(t)$ de amplitud o frecuencia más pequeñas necesitará un ancho de banda más pequeño que $B_{\text{máx}}$. Desde luego que este procedimiento desdeña el hecho de que la superposición no es aplicable a la modulación exponencial. Sin embargo, esta investigación de los espectros de tonos múltiples ha demostrado que los pares de bandas laterales de las frecuencias de batido están contenidos principalmente en el ancho de banda del tono dominante solo, (Fig. 6.8).

Por tanto, extrapolando la modulación de tono a una señal moduladora arbitraria, se define la *relación de desviación*†

$$\Delta \triangleq \frac{f_\Delta}{W} \tag{4}$$

como la desviación máxima dividida entre la frecuencia moduladora máxima, análogo al índice de modulación de la modulación de tono en el peor de los casos. Entonces, el ancho de banda de transmisión que se requiera para $x(t)$ es

$$B_T = 2M(\Delta)W \qquad M \geq 1 \tag{5}$$

donde se considera a Δ justo como a β para encontrar a $M(\Delta)$, digamos de la figura 6.10. A falta de curvas o tablas apropiadas de $M(\Delta)$, existen varias aproximaciones para B_T a las que se puede recurrir. Con los valores extremos de la relación de desviación, se encuentra que

$$B_T = \begin{cases} 2\Delta W = 2f_\Delta & \Delta \gg 1 \\ 2W & \Delta \ll 1 \end{cases}$$

de conformidad con los resultados obtenidos para la modulación de tono con β muy grande o muy pequeño. Ambas aproximaciones están combinadas en la relación conveniente

† Algunos autores emplean este término para f_Δ / f_c, lo cual no tiene el mismo significado.

$$B_T \approx 2(f_\Delta + W) = 2(\Delta + 1)W \qquad \begin{matrix} \Delta \gg 1 \\ \Delta \ll 1 \end{matrix} \tag{6}$$

conocida como *regla de Carlson*. En forma indebida, los sitemas de FM actuales tienen $2 < \Delta < 10$, por lo que la regla de Carson subestima en algo el ancho de banda de transmisión. Una aproximación mejor para el diseño de equipo es entonces

$$B_T \approx 2(f_\Delta + 2\,W) = 2(\Delta + 2)W \qquad \Delta > 2 \tag{7}$$

la cual, por ejemplo, se usaría para determinar los anchos de banda a 3 dB de los amplificadores de RF y FI.

Aplicando estas relaciones a la FM comercial, $f_\Delta = 75$ kHz y $W = 15$ kHz, por lo que $\Delta = 5$. Se ha encontrado ya que $M = 7$ para $\beta = 5$, por lo que la ecuación (5) da $B_T = 210$ kHz. (Las radios de FM de alta calidad tienen anchos de bandas en los pasos de FI de por lo menos 200 kHz.) La regla de Carlson subestima a B_T en aproximadamente 10%, dando $2(5 + 1) \times 15 = 180$ kHz, mientras que la ecuación (7) da justo en la marca con $2(5 + 2) \times 5 = 210$ kHz.

Se exhorta al lector a revisar estas aproximaciones y sus regiones de validez. En atención a la mayor parte de la literatura, se tomará con frecuencia a B_T como ocurre en la regla de Carlson, ecuación (6), teniendo en la mente sus limitaciones. Sin embargo, cuando $x(t)$ está lejos de ser de banda limitada, por ejemplo, un pulso rectangular, es necesario y prudente desarrollar el análisis espectral (si es posible) y determinar el ancho de banda a partir de éste.

EJERCICIO 6.3 Calcúlese B_T/W para $\Delta = 0.3$, 3 y 30 empleando las aproximaciones de las ecuaciones (6) y (7) donde sean aplicables. Repítase usando la ecuación (5) y la figura 6.10 y compárense los resultados.

FM de banda angosta (NBFM)

Las investigaciones sobre el ancho de banda, señalan la conclusión de que existen dos casos especiales de FM, correspondientes a valores muy pequeños o muy grandes de la relación de desviación. Estos casos tienen propiedades tan completamente diferentes, que se les han dado los nombres característicos de FM *de banda angosta* y FM *de banda ancha*, respectivamente.

La frecuencia modulada de banda angosta (NBFM) es, en muchas formas, similar a la modulación lineal de doble banda lateral. Para subrayar este punto, sea $\phi(t) = 2\pi f_\Delta \int_{-\infty}^{t} x(\lambda)d\lambda$ y supóngase que

$$|\phi(t)|_{\text{máx}} = \Delta \ll 1 \tag{8}$$

lo cual se toma como la condición para definir a la frecuencia modulada de banda angosta. Entonces, la onda modulada es

$$\begin{aligned} x_c(t) &= A_c \cos\,[\omega_c t + \phi(t)] \\ &= A_c[\cos\phi(t)\cos\omega_c t - \text{sen}\,\phi(t)\,\text{sen}\,\omega_c t] \\ &\approx A_c \cos\omega_c t - A_c\,\phi(t)\,\text{sen}\,\omega_c t \end{aligned} \tag{9a}$$

donde se han empleado las aproximaciones cos $\phi \approx$ y sen $\phi \approx \phi$. Puesto que la ecuación (9*a*) está en la forma de portadora en cuadratura y como

$$V_q(f) = \mathscr{F}[A_c\,\phi(t)] = A_c 2\pi f_\Delta \left[\frac{X(f)}{j2\pi f}\right]$$

la aplicación de la ecuación (4), (Sec. 5.1), da

$$X_c(f) \approx \frac{A_c}{2}[\delta(f - f_c) + \delta(f + f_c)] + \frac{A_c f_\Delta}{2}\left[\frac{X(f - f_c)}{f - f_c} - \frac{X(f + f_c)}{f + f_c}\right] \tag{9b}$$

el espectro como expresión que tiene la misma forma general del espectro de AM. Por tanto, si $x(t)$ es de banda limitada en W,

$$B_T \approx 2W \tag{10}$$

como se esperaba.

No debe sorprender que la frecuencia modulada de banda angosta no tenga una ventaja intrínseca sobre la modulación lineal más allá de lo que a producción de ruido se refiere. En consecuencia, la frecuencia modulada de banda angosta en raras ocasiones se usa para propósitos de transmisión aunque se le encuentra a menudo como un paso intermedio en la generación de FM de banda ancha.

FM de banda ancha (WBFM)

El carácter distintivo de la FM de banda ancha es que $\Delta \gg 1$, y en consecuencia B_T es grande en comparación con el ancho de banda del mensaje. En efecto B_T es *independiente* del ancho de banda del mensaje, es decir,

$$B_T \approx 2\Delta W = 2f_\Delta \tag{11}$$

puesto que $\Delta = f_\Delta/W$.

No existe una aproximación simple para la forma de onda de FM de banda angosta para $x_c(t)$, pero existe una aproximación para la *densidad espectral de potencia* $G_{xc}(f)$ en términos de la *función de densidad de probabilidad* $p_x(x)$ de la señal moduladora. La aproximación proviene de que, cuando $f_\Delta \gg W$, la desviación de frecuencia instantánea $f(t)$ varía tan lentamente que $X_c(t)$ se ve más o menos como una sinusoide ordenada de frecuencia $f_c + f(t)$ sobre intervalos de tiempo de orden $1/W$ lo cual es grande en comparación con $1/f_\Delta$. Por supuesto que $f(t)$ cambia puesto que $(t) = f_\Delta x(t)$, y, recurriendo a la ecuación (13), (Sec. 3.2).

$$p_f(f) = \frac{1}{f_\Delta}\,p_x\!\left(\frac{f}{f_\Delta}\right) \tag{12}$$

Por tanto, el tiempo fraccionario que $x_c(t)$ está en el intervalo de frecuencia $f - df$ a f es

$$\text{Prob}\,[f - df < f_c + f \leq f] = \text{Prob}\,[f - f_c - df < f \leq f - f_c] = p_f(f - f_c)\,df$$

por lo que $p_f(f - f_c)$ es proporcional a la densidad de potencia de frecuencia positiva en f. La densidad de potencia de doble lado es entonces

$$G_{x_c}(f) = \tfrac{1}{2}S_T p_f(f - f_c) + \tfrac{1}{2}S_T p_f(f + f_c)$$
$$= \frac{S_T}{2f_\Delta}\left[p_x\left(\frac{f - f_c}{f_\Delta}\right) + p_x\left(\frac{f + f_c}{f_\Delta}\right)\right] \tag{13}$$

donde $S_T = A_c^2/2$ es la potencia total.

Sujeta a la condición $\Delta \gg 1$, la ecuación (13) se puede emplear dondequiera que exista la función de densidad de probabilidad de $x(t)$, por lo menos en el sentido de la ocurrencia de la frecuencia relativa. El ejemplo y el ejercicio siguientes ilustran dos aplicaciones que incluyen una señal moduladora aleatoria y una no aleatoria, respectivamente.

Ejemplo 6.1 FM de banda ancha con modulación gaussiana ★

Sea $x(t)$ una señal aleatoria gaussiana de banda limitada en $W \ll f$ con $\bar{x} = 0$. No se puede aplicar la convención usual $|x(t)| \leq 1$ sino —recordando la desigualdad de Chebyshev— se requiere que

$$\sigma_x = 1/\kappa$$

tal que la probabilidad $[|x(t)| > 1] \leq 1/k^2$, lo cual se puede hacer pequeño en forma arbitraria eligiendo a k. De la ecuación (13), la densidad espectral de potencia es

$$G_{x_c}(f) = \frac{S_T}{2\sqrt{2\pi}\,\sigma_f}\left[e^{-(f-f_c)^2/2\sigma_f^2} + e^{-(f+f_c)^2/2\sigma_f^2}\right] \tag{14}$$

donde

$$\sigma_f = f_\Delta \sigma_x = \frac{f_\Delta}{\kappa}$$

como se bosqueja en la figura 6.11 para $f > 0$.

Claro que el ancho de banda no queda así bien definido; por un procedimiento similar al anterior, se determina a B_T teniendo en cuenta que se requiere que la potencia en el intervalo de frecuencia $f_c - B_T/2 \leq |f| \leq f_c + B_T/2$ sea igual a $(1 - 1/k^2)S_T$. Esto es equivalente a $B_T/2_{\sigma_f} = k$, y en consecuencia

$$B_T = 2\kappa\sigma_f = 2f_\Delta$$

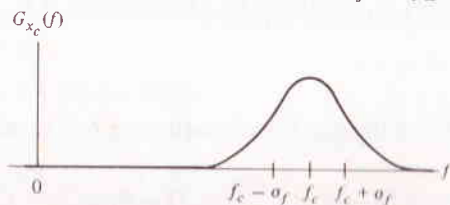



FIGURA 6.11
Espectro de potencia de FM de banda ancha (WBFM) con señal moduladora gaussiana, $f \leq 0$.

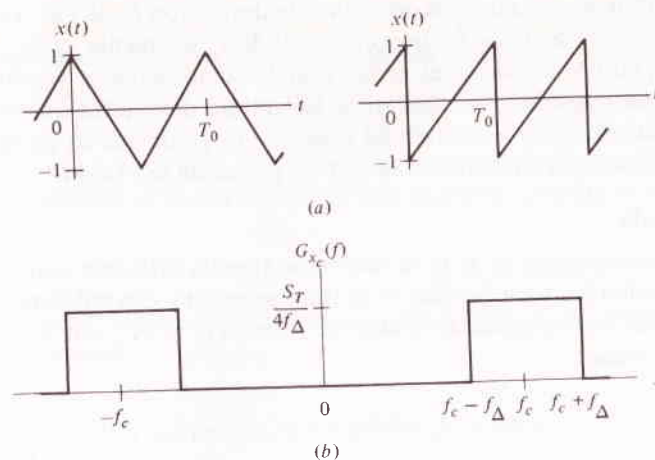



FIGURA 6.12

En el problema 6.15 se examina una definición alterna de ancho de banda.

EJERCICIO 5.4 Demuéstrese que ambas señales periódicas determinísticas de la figura 6.12*a* resultan ser $G_{x_c}(f)$ en la figura 6.12*b*, siempre y cuando $f_\Delta \gg 1/T_0$. (*Sugerencia*: Primero justifíquese que $p_x(x)$ está uniformemente distribuida.)

6.4 MODULACION DE FASE (PM)

Las modulaciones de frecuencia y de fase tienen muchas similitudes. En consecuencia, se abreviará bastante el análisis de la modulación de fase, aplicando en forma contundente los resultados obtenidos para la FM. No obstante, podría quizá notarse al principio una diferencia sutil pero significativa.

Recordando que la modulación de fase se definió como aquella que tiene la fase relativa $\phi(t) = \phi_\Delta x(t)$, donde ϕ_Δ es la constante de desviación de fase tal que

$$-\phi_\Delta \leq \phi(t) \leq \phi_\Delta$$

puesto que $|x(t)| \leq 1$. Considérese ahora la demodulación de una onda de modulación de fase: si se van a evitar las ambigüedades en la demodulación, $\phi(t)$ no debe exceder el intervalo $\pm\ 180°$; después de todo no existe diferencia física entre los ángulos de fase de $+\ 270°$ y $-\ 90°$, por ejemplo. En consecuencia, la constante de desviación está limitada por

$$\phi_\Delta \leq \pi \qquad \text{radianes} \tag{1}$$

Esta restricción es análoga en forma directa a la restricción $m \leq 1$ en AM, y puede llamarse de manera justa a ϕ_Δ el índice de modulación de fase.

No es necesario un límite semejante en la constante de desviación f_Δ de FM, ya que se puede distinguir a $f_c + f_\Delta$ de $f_c - f_\Delta$ siempre y cuando f_Δ sea menor que f_c. Por lo tanto, la desviación en FM se puede hacer tan grande como se desee, yendo hacia frecuencias portadoras mayores si es necesario. Lo último demuestra que el límite absoluto en ϕ_Δ, como contraste con el límite relativo en f_Δ, explica en parte por qué la superioridad del comportamiento de la FM en presencia del ruido.

Espectros y anchos de banda

Como lo anterior, el análisis espectral de la modulación de fase da principio con la modulación de tono. Sin embargo, para facilitar el asunto, se emplea una señal moduladora *senoidal* en vez de una cosenoidal. Entonces, con $x(t) = A_m \operatorname{sen} \omega_m t$, ecuación (5), (Sec. 6.1), se tiene

$$\begin{aligned} x_c(t) &= A_c \cos(\omega_c t + \phi_\Delta A_m \operatorname{sen}\omega_m t) \\ &= A_c \cos(\omega_c t + \beta_p \operatorname{sen}\omega_m t) \end{aligned} \qquad (2)$$

donde

$$\beta_p \triangleq A_m \phi_\Delta \qquad (3)$$

La razón para el empleo de la modulación de onda senoidal es obvia si se compara la ecuación (2) con la expresión para la FM, ecuación (2), (Sec. 6.2), ambas son idénticas salvo que β_p es independiente de la frecuencia de tono, mientras que β depende tanto de la amplitud del tono como de su frecuencia. Por tanto, los espectros de líneas de la modulación de fase tienen las mismas características generales que las de FM, con la siguiente excepción: Si se cambia la frecuencia moduladora f_m en tanto que se mantiene fija la amplitud A_m, β_p permanece constante, y sólo se altera el *espaciamiento* de las líneas. (Se verían afectados en el caso de la FM, tanto el espaciamiento de las líneas como la amplitud relativa de líneas.) Así, el lado izquierdo de la figura 6.5 se aplica tanto a la modulación de fase como a la modulación de frecuencia, mientras que el lado derecho no.

En relación a los anchos de banda, se podrían repetir los razonamientos de la Sec. 6.3 con modificaciones apropiadas. Sin embargo, sólo es necesario notar que $\Delta = f_\Delta / W$ es la *máxima desviación de fase* de una onda de FM bajo las peores condiciones de ancho de banda. Esto revela que Δ y ϕ_Δ son parámetros equivalentes, puesto que ϕ_Δ es la *máxima desviación de fase* de una onda de modulación de fase con $x(t)$ arbitraria está dado por

$$B_T = 2M(\phi_\Delta)W \qquad M \geq 1 \qquad (4a)$$

o

$$B_T \approx 2(\phi_\Delta + 1)W \qquad (4b)$$

lo cual es el equivalente aproximado a la regla de Carlson. Estas expresiones difieren del caso de la FM en que ϕ_Δ es independiente de W.

Modulación de fase de banda angosta (NBPM)

Dada la restricción $\phi_\Delta \leq 180°$, no existe modulación de fase de "banda ancha" en el mismo sentido que en la FM de banda ancha. Pero la modulación de fase de banda angosta, que tiene $\phi_\Delta \ll 180°$, es interesante y fácil de manejar desde el punto de vista analítico. Con desviación de fase pequeña, la onda de modulación de fase resulta ser

$$x_c(t) \approx A_c \cos\omega_c t - A_c \phi_\Delta x(t) \operatorname{sen}\omega_c t \qquad (5a)$$

como se desprende de la ecuación (9*a*), (Sec. 6.3), con $\phi(t) = \phi_\Delta x(t)$. El espectro es entonces

$$X_c(f) = \frac{A_c}{2}[\delta(f-f_c) + \delta(f+f_c)] + \frac{jA_c\phi_\Delta}{2}[X(f-f_c) - X(f+f_c)] \qquad (5b)$$

y $B_T = 2W$, como se esperaba.

Comparando el espectro de la modulación de fase de banda angosta con el de la AM, se demuestra que las similitudes aún son más pronunciadas que con la FM de banda angosta. Más aún, se aprecia en la ecuación (5) que la modulación de fase de banda angosta es un proceso de modulación *lineal*, al menos en forma aproximada, y se puede generar empleando dispositivos de modulación lineales. De manera concreta, un modulador balanceado con un desplazador de fase en cuadratura, pueden producir modulación de fase de banda angosta cuando se disponen como se muestra en el diagrama (figura 6.13)

Como la modulación de fase de banda angosta se genera con relativa facilidad, constituye el corazón de muchos sistemas de modulación de FM, o sea, un integrador que se agrega a la entrada para producir modulación de frecuencia en vez de modulación de fase. En la siguiente sección se describen ése y otros moduladores.

EJERCICIO 6.5 Demuéstrese que la fase en la salida de la figura 6.13 se da en forma más precisa por

$$\phi(t) = \phi_\Delta x(t) - \frac{\phi_\Delta^3}{3}x^3(t) + \frac{\phi_\Delta^5}{5}x^5(t) - \cdots \qquad (6)$$

por lo que existe distorsión en todas las armónicas impares a menos que ϕ_Δ sea muy pequeño.

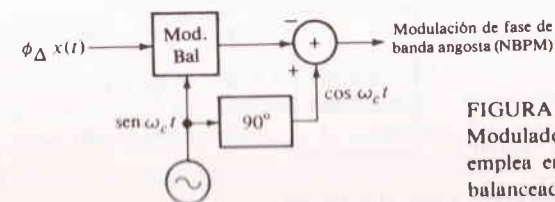



FIGURA 6.13
Modulador de fase de banda angosta que emplea en su construcción un modulador balanceado.

6.5 TRANSMISORES Y RECEPTORES

Cuando se analiza el equipo empleado para los sistemas de modulación exponencial, se debe tener presente que la frecuencia o fase instantáneas varían en forma lineal con la forma de onda del mensaje. Así, se necesitan dispositivos que produzcan o sean sensitivos a la variación de fase o de frecuencia en forma lineal. Se puede aproximar a tales características en muchas formas, pero algunas veces es difícil obtener convenientemente la relación lineal sobre un intervalo de operación amplio.

Por otra parte, la propiedad de amplitud constante de la modulación exponencial, es una ventaja definitiva desde el punto de vista componentes físicos. Por un detalle, el diseñador necesita no preocuparse por una disipación de potencia excesiva o por una ruptura de alto voltaje debidos a picos extremos en la forma de onda. Pero más importante, la distorsión de amplitud no lineal virtualmente no tiene efecto sobre la transmisión del mensaje, puesto que la información reside en los cruzamientos cero de la onda y no en la amplitud. (El corrimiento de fase o la distorsión por retardo, por supuesto, son intolerables.) También procede que cualesquier variaciones espurias de amplitud se pueden eliminar por medio de dispositivos recortadores de picos, conocidos como *limitadores*, sin afectar o remover el mensaje. En consecuencia, es posible una extensión considerable en el diseño y selección de equipo. Como ejemplo, los enlaces a repetidores en microondas para comunicaciones telefónicas de larga distancia, emplean principalmente la FM por que los amplificadores lineales de banda ancha que se necesitan para la modulación de amplitud no son útiles a las frecuencias de microondas.

Concretamente en la FM, existen dos métodos de generación básicos, conocidos como sistemas *directo* e *indirecto*. Se examinarán éstos a la vez y se abordará el tema de la demodulación de FM. Las presentaciones son extensas en la forma de diagrama a bloques; a los estudiantes interesados se les remite a la literatura para realizaciones de circuitos específicos y para análisis más detallados.

Modulación de frecuencia directa

De manera conceptual, la FM directa es precisamente directa y no requiere sino un *oscilador de voltaje controlado* cuya frecuencia de oscilación tiene una dependencia lineal respecto al voltaje aplicado. Esto se implementa con facilidad en la banda de microondas ($f_c \geq 1$ GHz), donde los dispositivos como la válvula *klystron* tienen, las características del oscilador de voltaje controlado lineal sobre un intervalo sustancial de frecuencia, en forma típica de varios megahertz. Si se desea una frecuencia portadora baja, la señal modulada se puede convertir a valores menores por heterodinación con la salida de un oscilador de frecuencia fija. Algunos generadores de prueba de laboratorio emplean precisamente esta técnica.

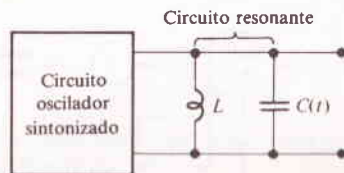



FIGURA 6.14
FM directa empleando reactancia variable.

† Clarke y Hess (1971, Caps. 11 y 12), etc.

De manera alterna, en el caso de frecuencias portadores bajas, es posible modular un oscilador de circuito sintonizado convencional, introduciendo un elemento de *reactancia variable* como parte del circuito resonante paralelo LC (Fig. 6.14). Si la capacidad tiene una dependencia del tiempo de la forma

$$C(t) = C_0 - Cx(t)$$

y si $Cx(t)$ es "lo bastante pequeño" y "lo bastante lenta", entonces la salida del oscilador es $A_c \cos \theta_c(t)$ donde

$$\frac{d\theta_c(t)}{dt} = \frac{1}{\sqrt{LC(t)}} = \frac{1}{\sqrt{LC_0}}\left[1 - \frac{C}{C_0}x(t)\right]^{-1/2} \tag{1}$$

Considerando a $\omega_c = 1/\sqrt{LC_0}$ y suponiendo que $|(C/C_0)x(t)| \ll 1$, el desarrollo en serie binominal de $d\theta_c(t)/dt \approx \omega_c[1 + (C/2C_0)\,x(t)]$, o

$$\theta_c(t) \approx 2\pi f_c t + 2\pi \frac{C}{2C_0} f_c \int_{-\infty}^{t} x(\lambda)\, d\lambda \tag{2}$$

lo cual es modulación de frecuencia con $f_\Delta = (C/2C_0)f_c$ —véase la ecuación (7), (Sec. 6.1). Puesto que $|x(t)| \leq 1$, la aproximación es buena y dentro del 1% cuando $C/C_0 > 0.013$ por lo que la desviación de frecuencia que se puede obtener está limitada por

$$f_\Delta = \frac{C}{2C_0} f_c \leq 0.006 f_c \tag{3}$$

Esta limitación cuantifica el significado de que $Cx(t)$ sea "pequeño" y es raro que imponga una dificultad de diseño. De manera similar, la condición usual $W \ll f_c$ asegura que $Cx(t)$ sea "lo bastante lento".

La reactancia variable se puede obtener de numerosas maneras. El caballo de batalla de los primero moduladores de FM directa, lo constituyó el tubo de reactancia, un pentodo polarizado de tal manera que su impedancia de salida contiene un término capacitivo proporcional al voltaje de rejilla. Otros moduladores emplean la capacitancia del efecto Miller, elementos a reactor saturable, o la reactancia de los diodos varactores.

La principal ventaja de la FM directa consiste en que son posibles grandes desviaciones de frecuencia sin operaciones adicionales. La principal desventaja reside en que la frecuencia portadora tiende a variar y debe estabilizarse por medio de un control de frecuencia de realimentación más elaborado. Como sólo hasta fechas muy recientes han surgido técnicas de estabilización satisfactorias, muchos transmisores de FM de los más viejos son del tipo indirecto.

MODULACIÓN EN FRECUENCIA INDIRECTA

El corazón de la FM indirecta es un *modulador de fase de banda angosta*, cuya frecuencia portadora es suministrada por medio de una fuente estable, por lo general un oscilador controlado a cristal, para asegurar la estabilidad. La figura 6.15 muestra las partes de un sistema indirecto completo como lo concibió Armstrong (1936). Las funciones son como sigue:

Antes de la modulación el mensaje se integra como para dar lugar a la modulación de frecuencia en lugar de a la de fase. El modulador de fase en sí mismo, puede ser como el que se muestra en la figura 6.13 o una variedad de otros tipos. En cualquier caso, la frecuencia modulada de banda angosta resultante a menudo contiene *distorsión* inherente (como se demostró en el ejercicio 6.5) a menos que la relación de desviación $\Delta_1 = f_{\Delta_1}/W$ sea muy pequeña. Por tanto, es necesario incrementar la relación de desviación después de la modulación, lo cual se hace por medio de una cadena de dobladores y triplicadores de frecuencia formando un *multiplicador de frecuencia*.

Un multiplicador de frecuencia ideal opera a frecuencia instantánea tal que si $f_{c1} + f_{\Delta_1}x(t)$ es la entrada, la multiplicación por n veces produce $f_{c2} + f_{\Delta_2}x(t) = nf_{c1} + nf_{\Delta_1}x(t)$. Nótese que éste es un proceso sutil, que afecta al *intervalo* de variación de frecuencia pero no al *régimen*. La multiplicación de una señal de tono modulado, por ejemplo, aumenta la frecuencia portadora y el índice de modulación, pero no a la frecuencia moduladora; así, las amplitudes relativas de las líneas de banda lateral se alteran, pero el espaciamiento de las líneas permanece igual.

En la figura 6.15, se escoge a n para dar la desviación final deseada, o sea, $n = f_\Delta/f_{\Delta_1} = \Delta/\Delta_1$. Pero este factor de multiplicación por lo general resulta en f_{c2} siendo mucho mayor que la frecuencia portadora deseada. La heterodinación de la salida del multiplicador con segundo oscilador controlado a cristal traslada el espectro *intacto* a la ubicación apropiada. (Por lo general, la heterodinación se hace a la mitad de la cadena de multiplicadores para impedir que las frecuencias centrales se hagan muy altas.) Todos estos pasos se hacen a bajos niveles de potencia, de manera que la componente final del sistema es un amplificador de potencia de RF de suficiente ancho de banda, pero cuya no linealidad no es decisiva.

A manera de ilustración con valores representativos, un transmisor indirecto de los primeros para FM comercial tenía $f_{c1} = 200$ kHz y $f_{\Delta_1} = 25$ Hz. Con $W = 15$

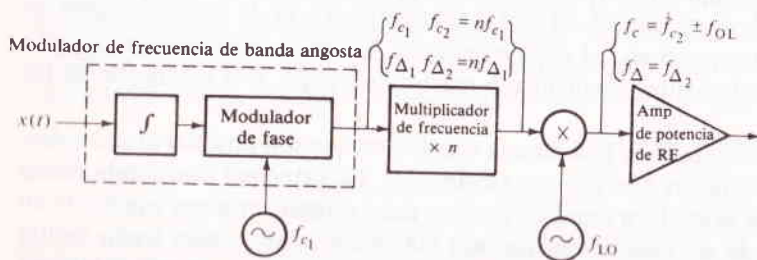



FIGURA 6.15
Un sistema de FM indirecta.

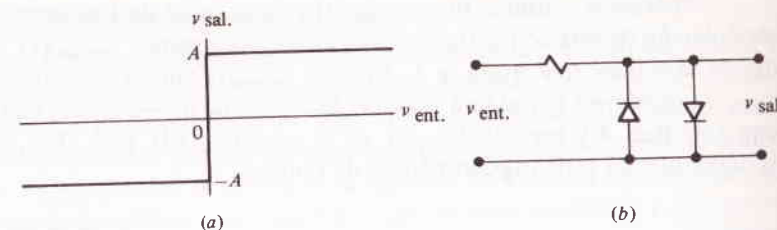



FIGURA 6.16
Limitador. (*a*) Características de transferencia ideal; (*b*) realización de circuito.

kHz, la relación de desviación inicial es $\Delta_1 = 25/(15 \times 10^3) \approx 2 \times 10^{-3}$, requiriéndose este valor para reducir al mínimo la distorsión y garantizar una buena fidelidad. Puesto que la desviación final va a ser $f_\Delta = 75$ kHz, el factor de multiplicación que se necesita es $n = 75 \times 10^3/25 = 3\,000$, el cual da lugar a seis triplicadores y dos dobladores. Después de la multiplicación, el espectro se localiza en $f_{c2} = 3\,000 \times 200$ kHz $= 600$ MHz, por lo que el segundo oscilador debe tener $f_{ol} \approx 600 \pm 100$ MHz para bajar a la portadora a la banda de FM de 88 a 108 MHz.

Estos cálculos indican que aunque la FM indirecta se obtiene alrededor del problema de la estabilidad de frecuencia, ello no es sin dificultades, principalmente la del factor de multiplicación. Además, los multiplicadores de banda ancha con el requisito de características de fase son casi tan complejos como los circuitos de control de frecuencia en FM directa, habiendo sido esto último mejorado en gran manera por el desarrollo de los contadores digitales. En consecuencia, existe un regreso a la FM directa con cantidades modestas de modulación de frecuencia para transmisores de desviación grande de alta calidad.

Limitadores

Un limitador es un dispositivo no lineal que remueve las variaciones de amplitud espurias de una onda modulada en forma exponencial sin destruir la modulación. Como tal, encuentra un lugar tanto en transmisores como en receptores de FM.

La figura 6.16*a* da la característica de transferencia de un *limitador riguroso* ideal, mientras que la figura 6.16*b* es una realización de circuito sencilla de ello. Puesto que el voltaje a través de un diodo semiconductor en sentido directo se aplana en aproximadamente 1 V, tan pronto como el diodo empieza a conducir en forma intensa, el voltaje de salida se recorta en forma rectangular cuando la amplitud de entrada sobrepasa ese nivel. Claro que si la entrada es una sinusoide no modulada, cuya amplitud es de varios volts, la salida será de manera esencial una onda cuadrada. En forma más general,

$$v_{sal.}(t) = A \operatorname{sgn} v_{ent.}(t) = \begin{cases} +A & v_{ent.}(t) > 0 \\ -A & v_{ent.}(t) < 0 \end{cases} \tag{4}$$

suponiendo un limitador riguroso ideal.

Supóngase entonces que $v_{\text{entrada}}(t)$ es una onda de frecuencia modulada o de modulación de fase con variaciones de amplitud digamos $v_{\text{entrada}}(t) = A_c(t)\cos\theta_c(t)$ donde se supone que $A_c(t) > 1$. Aunque $v_{\text{entrada}}(t)$ no es necesariamente periódica, v_{salida} se puede ver como una *función periódica de* θ_c —a saber, una onda cuadrada con amplitud A y periodo 2π. Así, de la ecuación (17), (Sec. 2.2), se puede escribir el desarrollo en serie trigonométrica de Fourier.

$$v_{\text{sal}} = \frac{4A}{\pi}\left[\cos\theta_c - \frac{1}{3}\cos 3\theta_c + \frac{1}{5}\cos 5\theta_c - \cdots\right] \tag{5}$$

Sustituyendo $\theta_c(t) = \omega_c t + \phi(t)$ se llega a

$$v_{\text{sal}}(t) = \frac{4A}{\pi}\left\{\cos\left[\omega_c t + \phi(t)\right] - \frac{1}{3}\cos\left[3\omega_c t + 3\phi(t)\right] + \cdots\right\} \tag{6}$$

lo cual es una onda cuadrada modulada en fase o en frecuencia.

Inspeccionando la ecuación (6) se demuestra que un filtro pasabanda de ancho de banda apropiado centrado en f_c extraerá el primer término, de una señal modulada de *amplitud constante*. Los limitadores que se construyen para usarse en receptores normalmente tienen tal filtro como una parte integral. Por otra parte, si el filtro pasabanda se centra en $3f_c$ o $5f_c$, se logra la multiplicación de frecuencia así como también limitación de amplitud, lo cual sería útil para transmisores. Obsérvese que el factor de multiplicación afecta tanto a la frecuencia portadora como a la fase relativa, como se examinó anteriormente.

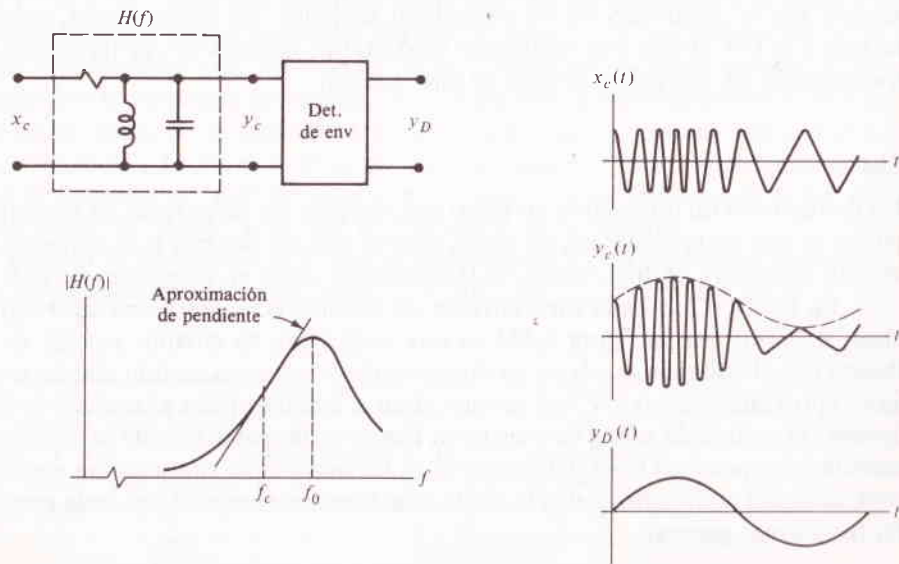



FIGURA 6.17
Detección de pendiente en FM, circuito y formas de onda.

Detección de FM discriminadores de frecuencia

Un demodulador de FM o *discriminador de frecuencia*, debe producir un voltaje de salida linealmente dependiente de la frecuencia de entrada. Ligeramente arriba o abajo de la frecuencia de resonancia, un circuito sintonizado sencillo más un detector de envolvente tienen esta propiedad sobre un intervalo limitado. Esto se conoce como *detección de pendiente* y se ilustra en la figura 6.17. Para un análisis cualitativo † de detección de pendiente, supóngase que $x_c(t)$ es tono modulado y f_c es menor que f_0. Entonces como $f(t)$ oscila arriba o abajo de f_c, la relación de amplitudes del circuito sintonizado convierte la variación de frecuencia a una variación de amplitud en la parte superior de la señal de FM, dando lugar a la forma de onda $y_c(t)$. Extrayendo sólo la variación de amplitud con un detector de envolvente (más componentes de CD) se produce $y_D(t)$ y se completa la demodulación.

Existen dos problemas con la detección de pendiente como se describe antes: el detector también responde a variaciones de amplitudes espurias de la entrada de FM y el intervalo de la pendiente lineal es muy pequeña. Un limitador que precede al circuito sintonizado cuida de lo primero, mientras que se puede lograr extender la linealidad empleando la configuración *balanceada* de la figura 6.18*a*. Un discriminador balanceado tiene dos circuitos resonantes, uno sintonizado sobre f_c y el otro por debajo. Así, conforme $f(t)$ cambia, las variaciones de amplitud están en direcciones opuestas, y tomando la diferencia de estas variaciones se tiene la característica voltaje a frecuencia de la figura 6.18*b*, la bien conocida curva *S*. La componente de CD se cancela en forma automática, sobrando la necesidad de un bloqueo de CD y con ello se mejora la respuesta a las frecuencias moduladoras bajas. Sin embargo, aún así, es necesario un limitador de entrada.

Los discriminadores balanceados se emplean a menudo como el elemento sensible a la frecuencia en control automático de frecuencia, lo cual tiene muchas aplicaciones en los sistemas de modulación exponencial. Ellos se adaptan con facilidad

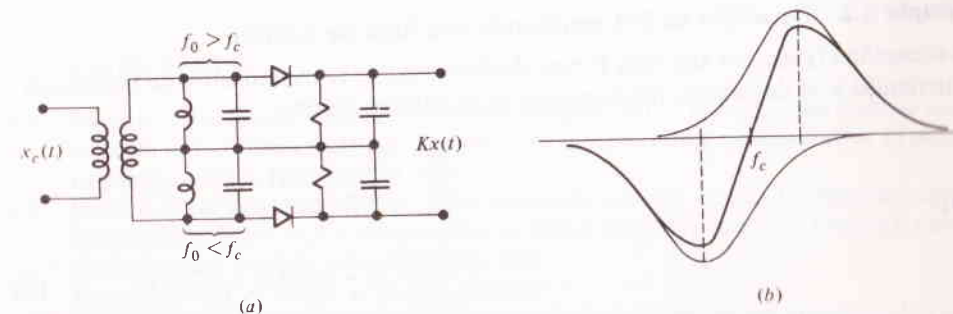



FIGURA 6.18
Discriminador balanceado. (*a*) circuito; (*b*) características frecuencias a voltaje.

† El análisis cuantitativo de la respuesta de circuito a la frecuencia variando en el tiempo es una cuestión difícil. Baghdady (1969, Cap. 19) presenta un examen conci[so del] problema y métodos de solución.

la banda de microondas, con cavidades resonantes que sirvan como circuitos sintonizados y diodos de cristal como detectores de envolvente. Otros demoduladores de FM comunes son el *discriminador de corrimiento de fase Foster-Seeley* y el *detector de relación*, cuyas descripciones se pueden encontrar en textos de radio y electrónica. El detector de relación es particularmente ingenioso y económico, ya que combina las operaciones de limitación y demodulación en una sola unidad. Para aplicaciones especializadas se pueden usar el de *sujeción de fase* (PLL), o el *lazo de sujeción de frecuencia (FLL);* en la sección 7.5 se examina un *pariente* de éstos, el receptor de FM de retroalimentación (FMFB).

Antes de dejar el tema de la demodulación, se debe puntualizar que cualquier dispositivo cuya salida sea la *derivada respecto del tiempo* de la entrada desarrollará una conversión de FM a AM y, por lo tanto, se puede emplear para detectar la FM. El que éste deba ser el caso procede del teorema de la diferenciación $dv/dt \longleftrightarrow j2\pi fV(t)$, de tal manera que un diferenciador tiene una característica frecuencia a amplitud linal. De manera específica, si $x_c(t) = A_c \cos \theta_c(t)$, entonces

$$\frac{dx_c(t)}{dt} = -A_c \frac{d\theta_c(t)}{dt} \operatorname{sen} \theta_c(t)$$

Pero, de la ecuación (7), (Sec. 6.1), $d\theta_c/dt = 2\pi[f_c + f_\Delta x(t)$, así

$$\frac{dx_c(t)}{dt} = -2\pi A_c[f_c + f_\Delta x(t)] \operatorname{sen} \theta_c(t) \qquad (7)$$

una forma de onda que tiene modulación en amplitud y en frecuencia. La figura 6.19 es un diagrama de un discriminador de frecuencia (con limitador) basado en la ecuación (7). A diferencia del discriminador balanceado, este sistema tendría una respuesta pobre a las bajas frecuencias.

Ejemplo 6.2 Detección de FM empleando una línea de retardo

La ecuación (7) sugiere aún otra forma de detección de la modulación de frecuencia recurriendo a la definición fundamental de la diferenciación,

$$\frac{dv(t)}{dt} = \lim_{\epsilon \to 0} \frac{1}{\epsilon} [v(t) - v(t - \epsilon)]$$

Así,

$$\frac{dx_c(t)}{dt} \approx \frac{1}{\tau} [x_c(t) - x_c(t - \tau)] \qquad (8)$$

$x_c(t)$ → Lim → $\frac{d}{dt}$ → Det de env. → Bloque de CD → $y(t)$

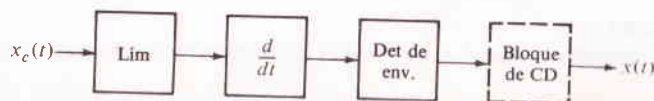



FIGURA 6.19
Discriminador de frecuencia que emplea conversión de FM a AM por diferenciación.

FIGURA 6.20
Aproximación a una línea de retardo para un diferenciador

De esta manera, la diferenciación se puede aproximar empleando una línea de retardo dispuesta conforme a la figura 6.20. La implementación no es particularmente difícil puesto que la aproximación requiere que τ sea pequeña en comparación con las variaciones de tiempo de $x_c(t)$, es decir, $\tau \ll 1/f_c$. Los detectores Foster-Seeley y de relación están basados en esta idea, pero emplean el corrimiento de fase más próximo al lineal de un circuito resonante para lograr el retardo en tiempo.

Receptores de FM

Muchos receptores de FM son de la variedad superheterodina. Difieren del de la figura 5.32 en dos aspectos: un limitador discriminador (o detector de relación) sustituye al detector de envolvente, y se incluye el control automático de frecuencia (AFC) para corregir la variación de frecuencia del oscilador local. Las radios comerciales de FM tienen un intervalo de sintonía de 88 a 108 MHz, $f_{FI} = 10.7$ MHz, y un ancho de banda en FI de 200 a 300 kHz. Así, el ancho de banda fraccionario del amplificador de FI es aproximadamente de 2×10^{-2}, el mismo que en los superheterodinos de AM.

Existe una tercera diferencia entre los receptores de FM y los de AM, la que en un receptor de FM se puede tener un *filtro de desacentuación* a la salida. El objeto de este filtro se explicará cuando se traten la interferencia y el ruido en los sistemas de modulación exponencial.

6.6 PROBLEMAS

6.1 (Secc. 6.1) Bosquéjese y rotúlese en forma completa $\phi(t)$ y $f(t)$ para una onda de modulación de fase y de modulación de frecuencia cuando la señal moduladora es un pulso triangular, $x(t) = A\Delta(t/\tau)$.

6.2 (Secc. 6.1) Un generador de barrido de frecuencia produce una salida sinusoidal cuya frecuencia es f_1 en $t = 0$ y se incrementa en forma lineal a f_2 en $t = T$. Obténgase una expresión para la salida en la forma $\cos \theta(t)$.
Respuesta: $\theta(t) = 2\pi[f_1 t + (f_2 - f_1)t^2/2T]$.

6.3 (Secc. 6.1) Otras dos formas posibles de modulación exponencial son la modulación integral de fase donde $\phi(t) = K dx(t)/dt$, y la modulación de aceleración de fase, donde $f(t) = K\int^t x(\lambda)\,d\lambda$. Agréguense éstas a la tabla 6.1 y encuéntrese los máximos valores de $\phi(t)$ y $f(t)$ para todos los cuatro tipos cuando $x(t) = \cos 2\pi f_m t$.

6.4 (Secc. 6.2) Empleando las ecuaciones (5) y (6), desarróllense todos los detalles entre las ecuaciones (4) y (7*a*).

6.5 (Secc. 6.2) Constrúyanse los diagramas fasoriales para FM de tono modulado con $\beta = 1.0$ cuando $\omega_m t = 0$, $\pi/4$, y $\pi/2$. Inclúyanse por lo menos tres pares de bandas laterales, y verifíquese que el corrimiento de fase resultante es correcto y la amplitud es constante.

6.6 (Secc. 6.2) Una señal de FM de tono modulado con $\beta = 1.0$ y $f_m = 200$ pasa a través de un filtro pasabajas ideal con $B = 500$ centrado en f_c. Dibújese el espectro de líneas y el diagrama fasorial de la señal de salida y descríbase qué se observaría si la forma de onda se exhibe en un osciloscopio.

6.7 (Secc. 6.2) La componente frecuencia portadora de una señal de FM de tono modulado es cero dondequiera que $J_0(\beta) = 0$. Empleando esta propiedad, idéese un procedimiento experimental para calibrar las características de desviación de un modulador de FM.

6.8⋆ (Secc. 6.2) Aplíquese la ecuación (10) al caso donde $x(t)$ es una onda cuadrada de amplitud unitaria. Escríbase $x_c(t)$ como una suma de sinusoides, y bosquéjese el espectro de amplitud tomando $f_\Delta \gg 1/T_0$.

6.9 (Secc. 6.3) Un mensaje tiene $W = 10$ kHz. Encuéntrese el ancho de banda de transmisión de FM cuando $f_\Delta = 0.1$, 1.0, 10, 100, y $1\,000$. En cada caso selecciónese la relación apropiada de las ecuaciones (5), (6) y (7).

6.10 (Secc. 6.3) Un sistema de FM tiene $f_\Delta = 10$ kHz. Calcúlese B_T para:
(*a*) Transmisión de voz pobremente inteligible.
(*b*) Transmisión de voz en telefonía de alta calidad.
(*c*) Transmisión de audio de alta fidelidad.
(*Sugerencia*: Empléese la tabla 4.2 y escójase la curva apropiada en la figura 6.10 para cada caso.)

6.11 (Secc. 6.3) Una señal analógica de datos con $W = 2$ kHz va a ser enviada por medio de FM en un sistema con $B_T = 12$ kHz. Si se desea la reproducción en alta fidelidad, ¿cuál es el valor máximo permitido de f_Δ?
Respuesta: 1.9 kHz.

6.12⋆ (Secc. 6.3) Ciertos sistemas de telemetría de multicanalización por división en frecuencia emplean la modulación de subportadora de FM de *ancho de banda proporcional* en la cual los parámetros del canal *i*-ésimo se relacionan por $f_{\Delta i} = af_{c_i} = W_i/b$, donde a y b son constantes.
(*a*) Demuéstrese que B_{T_i}, el ancho de banda de la señal subportadora *i*-ésima, es proporcional a f_{c_i}, suponiendo que es aplicable la regla de Carlson.
(*b*) Obténgase un límite inferior en f_{c_i} en términos de f_{c_1}, a, y b.
(*Sugerencia*: Para evitar el traslape de bandas laterales, es necesario que $f_{c_i} - B_{T_i}/2 \geq f_{c_{i-1}} + B_{T_{i-1}}/2$.)

6.13 (Secc. 6.13) Dedúzcase la ecuación (9*b*) a partir de la (9*a*).

6.14 (Secc. 6.3) Si Θ es un ángulo aleatorio uniformemente distribuido sobre $2p$ radianes y $X = \cos\Theta$, se puede demostrar que $p_x(x) = [\pi\sqrt{A^2 - x^2}]^{-1}\Pi(x/2A)$. Empléese este hecho y la ecuación (13) para encontrar $G_{x_c}(f)$ cuando $x(t) = \cos 2\pi f_m t$, $f_m \ll f_\Delta$. Bosquéjese $G_{x_c}(f)$ para $f > 0$ y compárese con la figura 6.5.

6.15⋆ (Secc. 6.3) El *ancho de banda rms* de una señal pasabanda centrada en f_c está definido por $B_{\text{rms}}^2 = (8/S_T)\int_0^\infty (f - f_c)^2 G_{x_c}(f)\,df$.
(*a*) Suponiendo que se aplica la ecuación (13), demuéstrese que ésta reduce a $B_{\text{rms}} = 2f_\Delta\sigma_x$ cuando $f_c \gg f_\Delta$ y $\bar{x} = 0$.
(*b*) Encuéntrese B_{rms} para una señal moduladora gaussiana normalizada tal que $[|x(t)| > 1] = 0.2$.

6.16 (Secc. 6.4) Se aplica modulación de tono en forma simultánea a un modulador de FM y de PM y los espectros de líneas de salida son idénticos. ¿Cómo cambiarán estos espectros, en general, si: la frecuencia del tono aumenta o disminuye; la amplitud del tono aumenta o disminuye?

6.17 (Secc. 6.4) Si $W = 10$ kHz. En forma aproximada ¿qué porcentaje se ocupa de B_T cuando la señal moduladora es un tono de amplitud unitaria es $f_m = 0.1$, 1.0 o 5.0 kHz y la modulación es FM con $f_\Delta = 30$ kHz? Repítase para modulación de fase con $\phi_\Delta = 3$ y compárese.

6.18 (Secc. 6.4) Analícense las propiedades de la modulación integral de fase y de la modulación de aceleración de fase definidas en el problema 6.3, concediendo una atención particular a sus espectros de líneas y requisitos de ancho de banda.

6.19 (Secc. 6.5) Cuando una sinusoide de frecuencia f_c se aplica a un sistema lineal con una componente variando con el tiempo cuyas variaciones son "lentas" en comparación con f_c ∈ se puede emplear un análisis de CA de estado *cuasiestacionario* para aproximar la salida escribiendo la función de transferencia como $H(f_c, t)$, donde la dependencia del tiempo refleja a la componente variando en el tiempo. Empléese esta aproximación para analizar el modulador de fase RC de la figura P6.1 donde $R(t) = R_0 - R_x(t)$ con $R \ll R_0$ y $\omega_c R_0 C \gg 1$. Estímese la máxima desviación de fase que se puede lograr sin distorsión significativa en $x_c(t)$.

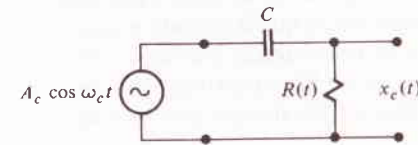



FIGURA P6.1

6.20 (Secc. 6.5) La figura P6.2 es un generador de FM directa que emplea un *oscilador controlado a voltaje*. La salida del oscilador tiene $d\theta(t)/dt = 2\pi[f_0 + K_1 v(t) + \epsilon(t)]$, donde $v(t)$ es el voltaje de entrada (de control) y $\epsilon(t)$ es una variación de frecuencia aleatoria. Para evitar esta variación, la salida se mezcla con un oscilador estable y realimenta a través de un discriminador cuyo voltaje de salida es $K_2[f_{\text{entrada}}(t) - f_{\text{FI}}]$. Suponiendo que $x(t)$ tiene $\bar{x} = 0$ y es de banda limitada en W, mientras que $\epsilon(t)$ es de banda limitada en W_ϵ, demuéstrese que esta disposición reduce la variación en un factor de $1/(1 + K_1K_2)$. ¿Por qué es necesario que $\bar{x} = 0$?

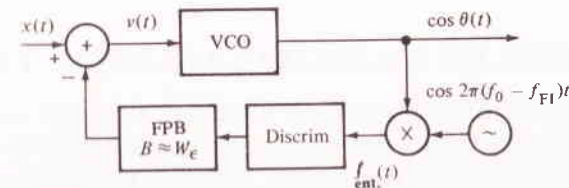



FIGURA P6.2

6.21 (Secc. 6.5) Se agrega un integrador a la figura 6.13 tal que la entrada al modulador balanceado resulta ser $\phi_\Delta \int^t x(\lambda)\,d\lambda$, $\phi_\Delta \ll \pi$.

(*a*) Demuéstrese que la salida tiene $f(t) \approx f_\Delta x(t)\{1 - [\phi_\Delta \int^t x(\lambda)d\lambda]^2\}$, donde $f_\Delta = \phi_\Delta / 2\pi$.
(*b*) Tomando a $x(t) = \cos 2\pi f_m t$, demuéstrese que $f(t)$ tiene distorsión de tercera armónica con una amplitud relativa de aproximadamente $(f_\Delta/2f_m)^2$.
(*c*) Explíquese brevemente por qué la frecuencia moduladora *más baja* sufre la mayor distorsión.

6.22 (Secc. 6.5) Una señal de banda base multicanalizada cuyo espectro cubre de 100 a 2 500 kHz va a ser transmitida con modulación de portadora de FM. Haciendo uso de la información del problema 6.21, diséñese un transmisor de FM indirecta con las siguientes especificaciones: f_c = 10 kHz, Δ = 2, distorsión de tercera armónica menor del 10%.

6.23★ (Secc. 6.5) Se va a diseñar un transmisor de FM indirecta para f_c = 10 MHz y f_Δ = 12 kHz. El mensaje tiene W = 4 kHz. La cadena de multiplicadores consistirá de seis triplicadores, siendo cada uno un limitador riguroso seguido de un filtro pasabanda. Cada filtro debe satisfacer el ancho de banda fraccionario obligado de la ecuación (15), (Sec. 5.1), pero sus anchos de banda deben ser mayores que el ancho de banda de la señal en cualquier punto, mientras que las señales no deseadas se rechazan. Dibújese un diagrama completo del transmisor, especificando valores para $f_{\Delta 1}$, f_{c1}, y f_{OL}, y anchos de banda y frecuencias centrales de los filtros pasabanda. Colóquese el mezclador oscilador local tal que la frecuencia no exceda de 12 MHz.

6.24 (Secc. 6.5) La aproximación de CA de estado cuasiestacionario mencionada en el problema 6.19 se puede modificar para encontrar la respuesta de un filtro a una señal de entrada de FM que tiene $|f(t)| \ll f_c$. Se ve en forma simple que la entrada es una sinusoide de frecuencia $f_c + f(t)$ tal que la función de transferencia resulta ser $H[f_c + f(t)]$. Empleando esta aproximación, demuéstrese que un filtro pasabajas RC con $B = f_c$ puede reemplazar al circuito sintonizado de la figura 6.17, siempre que sea aceptable la inversión de signo de la señal moduladora.

6.25★ (Secc. 6.5) El circuito sintonizado de la figura 6.17 tiene $H(f) \approx [1 + j(2Q/f_0)(f - f_0)]^{-1}$ para $f > 0$, donde $Q = R\sqrt{C/L} \gg 1$ y $f_0 = 1/2\pi\sqrt{LC}$.
(*a*) Demuéstrese que la pendiente de $H(f)$ es máxima en $f_0 - f_0/2\sqrt{2Q}$.
(*b*) Empéese el método del problema 6.24 para aproximar $y_c(t)$, suponiendo que f_c esta en el punto de máxima pendiente.

6.26★ (Secc. 6.5) Idéese y analícese un sistema de control automático de frecuencia (AFC) para un receptor superheterodino de FM que emplea un oscilador controlado a voltaje, problema 6.20 que tiene $f_0 = f_c - f_{FI}$ en lugar del oscilador local.

7

EL RUIDO EN LA MODULACIÓN DE ONDA CONTINUA (CW)

El análisis de las dos clases de modulación de onda continua, prosigue en este capítulo y cubre el comportamiento de los sistemas de comunicación de esta clase de onda ante ruido aditivo. Así, se enfocará ahora a la transmisión *pasabanda* comparándola con la transmisión de banda base. Sin embargo, muchos de los conceptos tratados (Cap. 4), se aplicarán al caso pasabanda.

Estos temas incluyen modelos y parámetros de sistema, interferencia, ruido pasabanda y relaciones señal a ruido. El capítulo concluye con una comparación de los tipos de modulación lineal y exponencial entre sí y la transmisión de banda base.

7.1 MODELOS Y PARÁMETROS DEL SISTEMA

La figura 7.1 muestra el diagrama de la forma general de un sistema de comunicación de onda continua. Como antes, el mensaje $x(t)$ se considera una señal analógica de ancho de banda W con $x(t) \leq 1$ y $\overline{x^2} \leq 1$. Se supone que el transmisor es ideal y el canal proporciona una transmisión cercana a aquélla sin distorsión † sobre

† De otra manera, se requeriría igualar. Por lo general, tales igualadores deben ser unidades pasabanda localizadas antes del detector, puesto que la detección es una operación no lineal. Una excepción es el caso de la detección sincrónica, pues se puede emplear la igualación de banda base.

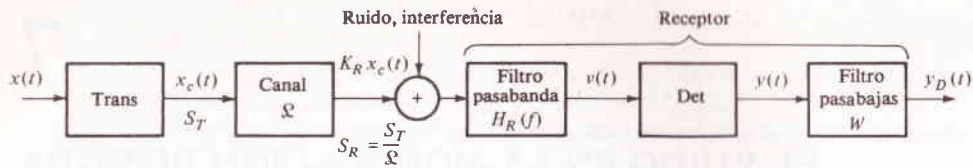



FIGURA 7.1
Sistema de comunicación de CW (onda continua).

el ancho de banda de transmisión B_T, con pérdida de potencia $\mathscr{L}$ y retardo en tiempo despreciable. Se supone además que la porción de predetección del receptor se puede modelar como un filtro pasabanda $H_R(f)$ que tiene una ganancia unitaria sobre $B_R \geq B_T$. En un receptor superheterodino, $H_R(f)$ es la respuesta de frecuencia del amplificador de FI referido (trasladado) a la frecuencia portadora (véase Fig. 5.33*d*).

Bajo estas condiciones, y descartando por el momento cualquier interferencia o ruido aditivo, la señal en la entrada del detector es

$$v(t) = K_R x_c(t) \tag{1}$$

donde

$$K_R \triangleq \frac{1}{\sqrt{\mathscr{L}}} \tag{2}$$

En forma similar, $A_R = K_R A_c$ es la amplitud de la portadora en la entrada del detector. Entonces, en términos de la potencia transmitida S_T,

$$\langle v^2(t) \rangle = K_R{}^2 \langle x_c{}^2(t) \rangle = \frac{S_T}{\mathscr{L}} = S_R \tag{3}$$

siendo S_R la potencia recibida. Las expresiones para $x_c(t)$, S_T y B_T se han desarrollado en los capítulos 5 y 6 para los diferentes tipos de modulación.

Aunque se puede transmitir una onda limpia, en forma virtual libre de ruido, la señal entregada al demodulador va siempre acompañada de ruido, incluso el que se genera en las etapas precedentes del receptor mismo. Además, puede haber señales *interferentes* en la banda deseada que no sean rechazadas por $H_R(f)$. Tanto el ruido como la interferencia son componentes indeseables en la salida del detector y cuando se manifiestan, se escribirá la señal contaminada $v(t)$ en la forma de envolvente y fase o de portadora en cuadratura, o sea,

$$v(t) = R_v(t) \cos [\omega_c t + \phi_v(t)] \tag{4a}$$

$$= v_i(t) \cos \omega_c t - v_q(t) \operatorname{sen} \omega_c t \tag{4b}$$

lo cual facilita el análisis de la señal demodulada $y(t)$. De manera específica, la operación de demodulación se representa por medio de los siguientes modelos matemáticos idealizados: †

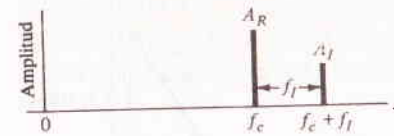



FIGURA 7.2
Espectro de líneas para sinusoides interferentes.

$$y(t) = \begin{cases} v_i(t) & \text{Detector sincrónico} \quad (5a) \\ R_v(t) - \overline{R}_v & \text{Detector de envolvente} \quad (5b) \\ \phi_v(t) & \text{Detector de fase} \quad (5c) \\ \dfrac{1}{2\pi}\dfrac{d\phi_v(t)}{dt} & \text{Detector de frecuencia} \quad (5d) \end{cases}$$

El término $\overline{R}_v = \langle R_v(t) \rangle$ en la ecuación (5*b*) refleja que el bloque de CD se encuentra de manera normal en un detector de envolvente; los otros detectores pueden tener o no esta característica. Se podría incluir una constante de detección en la ecuación 5 pero no agregaría algo en favor de la generalidad.

Nótese, sin embargo, que $y(t)$ no es necesariamente igual a la señal de salida final $y_D(t)$ en el destino, puesto que en la figura 7.1 se tiene un filtro pasabajas a continuación del detector. Por lo tanto, suponiendo que el filtro pasabajas solo remueve a cualesquier componentes de frecuencia fuera de banda,

$$y_D(t) = \int_{-W}^{W} Y(f) e^{j\omega t}\, df \tag{6}$$

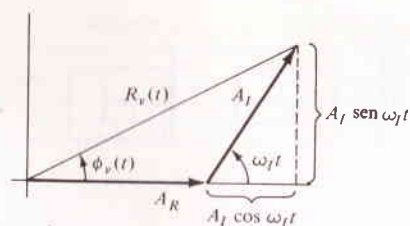
donde $Y(f) = \mathscr{F}[y(t)]$. A esta operación se le llama filtraje *postdetección*, para diferenciarla del filtraje de *predetección* pasabanda desarrollado por $H_R(f)$.

Por último, cuando se llega al ruido aditivo, se supondrá que es ruido blanco de valor medio cero con densidad espectral $G(f) = \eta/2$ en la entrada a $H_R(f)$. Así, existe ruido *pasabanda* aditivo en la entrada del detector. En la sección 7.3 se desarrolla la representación del ruido pasabanda que será empleado en las subsecuentes secciones 7.4 y 7.5 para encontrar a $(S/N)_D$, o sea, la relación señal a ruido en el destino. Al analizar la interferencia, se propicia el acceso a esos temas.

7.2 INTERFERENCIA

Se principia por considerar un caso muy sencillo, el de una portadora no modulada con una onda cosenoidal interferente (Fig. 7.2). La señal interferente tiene una amplitud A_I y frecuencia $f_c + f_I$. La señal total que entra al demodulador es la suma de dos sinusoides

† Idealizados en el sentido de una sincronización perfecta, de una limitación de amplitud perfecta etc.

FIGURA 7.3
Diagrama fasorial para sinusoides interferentes.

$$v(t) = A_R \cos \omega_c t + A_I \cos (\omega_c + \omega_I)t$$

así

$$R_v(t) = \sqrt{(A_R + A_I \cos \omega_I t)^2 + (A_I \operatorname{sen} \omega_I t)^2}$$

$$\phi_v(t) = \arctan \frac{A_I \operatorname{sen} \omega_I t}{A_R + A_I \cos \omega_I t} \tag{1}$$

como se desprende de la construcción fasorial de la figura 7.3.

Para valores arbitrarios de A_R y A_I, estas expresiones no se pueden simplificar más. Sin embargo, si la interferencia es pequeña en comparación con la portadora, el diagrama fasorial muestra que la envolvente resultante es esencialmente la suma de las componentes en fase, mientras que la componente en cuadratura determina el ángulo de fase. O sea, que si $A_I \ll A_R$, entonces

$$R_v(t) \approx A_R + A_I \cos \omega_I t \tag{2a}$$

$$\phi_v(t) \approx \frac{A_I}{A_R} \operatorname{sen} \omega_I t \tag{2b}$$

y en consecuencia

$$v(t) = A_R(1 + m_I \cos \omega_I t) \cos (\omega_c t + m_I \operatorname{sen} \omega_I t) \tag{3}$$

donde

$$m_I \triangleq \frac{A_I}{A_R} \ll 1$$

Se obtiene el mismo resultado de los desarrollos de primer orden de la ecuación (1).

Por otro lado, si $A_I \gg A_R$, el análisis se realiza considerando la interferencia como referencia y descomponiendo el fasor portadora, lo cual da $v(t) = A_I(1 + m_I^{-1} \cos \omega_I t)\cos[(\omega_c + \omega_I)t - m_I^{-1} \operatorname{sen} \omega_I t]$, como se podría esperar.

Se ve en la ecuación 3 que la onda interferente *modula la amplitud y la fase* de una portadora precisamente como un tono modulante f_I con índice de modulación m_I. Por otra parte, con interferencia intensa, se puede considerar que la portadora es modulada por la onda interferente. En uno u otro caso, la frecuencia de modulación aparente es la diferencia de frecuencia f_I.

EJERCICIO 7.1 Considérese el caso de la interferencia debida a reflexiones de un avión volando a baja altura que $A_I \approx A_R$ y $f_I = vf_c/c$ donde v es la velocidad relativa y c la velocidad de la luz. Demuéstrese que $R_v(f) = 2A_R\ |\cos(\pi v f_c t/c)|$.

Interferencia en la modulación lineal

Supóngase que existe una pequeña interferencia de amplitud en un sistema de AM con detección de envolvente. Empleando la ecuación (2*a*) de esta sección y las ecuaciones (5*b*) y (6), (Sec. 7.1), la señal de salida resulta ser

$$y_D(t) = \begin{cases} A_I \cos \omega_I t & |f_I| < W \\ 0 & |f_I| > W \end{cases} \tag{4}$$

puesto que $\overline{R}_v = A_R$. De manera similar, para detección sincrónica,

$$y_D(t) = A_R + A_I \cos \omega_I t \qquad |f_I| < W \tag{5}$$

puesto que $v_i(t) = A_R + A_I \cos \omega_I t$—(véase Fig. 7.3). La componente de CD de la ecuación (5) puede o no estar bloqueada. En uno u otro caso, cualquier interferencia en la banda $f_c \pm W$ produce una señal detectada cuya amplitud depende sólo de A_I, o sea la amplitud de la interferencia, siempre y cuando $A_I \ll A_R$.

El lector puede preguntarse por qué puede existir confusión en este problema sencillo cuyos resultados son casi obvios a raíz de una inspección. La razón es doble: primera, como se podría inferir de la ecuación (1), las complicaciones se multiplican en forma rápida cuando la interferencia es grande o la portadora es modulada; segundo, ocurren algunas diferencias interesantes y significativas en la modulación exponencial.

Interferencia en la modulación exponencial

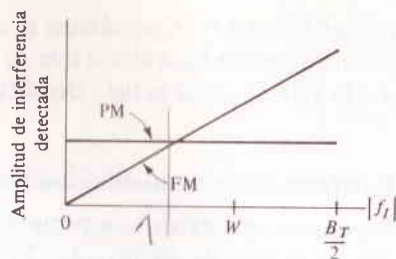
Con un detector de fase o frecuencia, se encuentra la interferencia insertando la ecuación (2*b*) de esta sección en las ecuaciones (5*c*) y (5*d*), (Sec. 7.1). Así, para f_I

$$y_D(t) = \frac{A_I}{A_R} \operatorname{sen} \omega_I t \qquad \text{PM} \tag{6}$$

$$y_D(t) = \frac{A_I f_I}{A_R} \operatorname{sen} \omega_I t \qquad \text{FM} \tag{7}$$

$< W$, donde f_I aparece como un factor de multiplicación en la ecuación (7) pero no en la (6) en razón de la diferenciación de $\phi_v(t)$.

Comparando las ecuaciones (6) y (7) con las (4) y (5), y el hecho de que $A_I/A_R \ll 1$, se encuentra que la modulación exponencial es menos vulnerable a la interferencia de amplitud pequeña que la modulación lineal, siendo iguales todos los demás factores. Más aún, de las ecuaciones (6) y (7) se desprende que la FM es menos vul

FIGURA 7.4
Amplitud de interferencia detectada, como función de $|f_I|$ para una onda interferente de frecuencia $f_c + f_I$.

nerable que la PM (modulación de fase) cuando $|f_I|$ es pequeña, puesto que la interferencia detectada es proporcional tanto a la amplitud como a la frecuencia de la onda interferente. En los sistemas de PM, como en la modulación lineal, sólo la amplitud entra en escena.

Esta última diferencia se puede entender con la ayuda de consideraciones físicas sencillas. La intensidad de una señal detectada en FM depende de la máxima *desviación de frecuencia*. Las ondas interferentes cercanas a la frecuencia portadora no pueden provocar un cambio significativo en la frecuencia de la resultante y, por lo tanto, producen poco efecto. A mayor diferencia entre f_c y $f_c + f_I$, corresponde una mayor desviación de frecuencia que hace que la salida demodulada sea proporcional a $|f_I|$. Pero para PM, la máxima *desviación de fase* depende sólo de las amplitudes relativas, como se muestra por medio del diagrama fasorial de la figura 7.3.

El comportamiento de la FM y de la PM respecto a la interferencia, se muestra mejor construyendo la gráfica de la amplitud de la señal $y(t)$ *no filtrada* como una función de $|f_I|$ (Fig. 7.4). Por supuesto, $y_D(t) = y(t)$ para $|f_I| < W$, pero $y(t)$ incluye frecuencias por lo menos arriba de $B_T/2$ puesto que $H_R(f)$ permite el paso de $f_c \pm B_T/2$ (suponiendo que $B_R \approx B_T$) y $B_T/2$ sobrepasa a W con la modulación exponencial. La figura 7.4 ilustra que cuando la interferencia se debe a una estación cocanal, † entonces $f_c + f_I \approx f_c$, si $|f_I|$ es pequeña, y es obvio que la FM es menos vulnerable. A la inversa, la PM tiene mejor comportamiento con respecto a la interferencia de *canal adyacente*, en la cual $|f_I|$ es relativamente grande. En uno u otro caso, debe seguir al demodulador un filtro pasabajas de ancho de banda W para eliminar a las componentes de interferencia detectada que están fuera de la banda del mensaje, pero que no son rechazados por el filtro de predetección, es decir, interferencia en $W < |f_I| < B_T/2$. Tal filtraje de postdetección es deseable pero no constituye una necesidad en los sistemas de modulación lineal, dado que el ancho de banda de transmisión no es mayor que $2W$.

Filtraje de desacentuación y de preacentuación

El hecho de que la interferencia de FM detectada sea más severa a valores grandes de $|f_I|$ sugiere un método para mejorar el comportamiento del sistema con un filtraje selectivo de postdetección, llamado *filtraje de desacentuación*.

† Las expresiones *co-canal* (*canal propio*) y *canal adyacente*, se refieren a las estaciones que tienen asignada la misma frecuencia portadora o una adyacente. La interferencia co-canal también procede de frecuencias imágenes.

FIGURA 7.5
El demodulador de FM completo.

Supóngase que al demodulador le sigue un filtro pasabajas que tiene una relación de amplitudes que empieza a decrecer en forma gradual *debajo* de W; esto *desacentuará* la porción de alta frecuencia del mensaje y con ello se reducirá la interferencia más seria. Se requiere aún un filtro pasabajas de corte agudo (ideal) para remover cualquier componente residual que esté encima de W, por ello, el demodulador completo está formado de un detector de frecuencia, un filtro de desacentuación y un filtro pasabajas, figura 7.5.

Obviamente, el filtraje de desacentuación también atenúa las componentes de alta frecuencia del mensaje mismo, ocasionando distorsión de la señal de salida, a menos que se tomen medidas correctivas. Pero es sencillo compensar la distorsión por desacentuación *predistorsionando* o *preacentuando* la señal moduladora en el transmisor antes de la modulación. Las características de preacentuación y desacentuación del filtro deben estar relacionadas por

$$H_{pe}(f) = \frac{1}{H_{de}(f)} \qquad |f| \leq W \tag{8}$$

para lograr una transmisión sin distorsión, neta.

En esencia, se preacentúa el mensaje antes de la modulación (donde no exista la interferencia) para poder desacentuar la interferencia relativa al mensaje después de la demodulación. La técnica es por medios que no restringen a los sistemas de FM; se puede emplear con ventaja dondequiera que predominen las contaminaciones no deseadas, en ciertas porciones de la banda del mensaje. Por ejemplo, las altas frecuencias se preacentúan en forma común en grabación de sonido de manera que el ruido de superficie de alta frecuencia se puede desacentuar durante la reproducción.† Por otra parte, los sistemas de modulación de fase y lineal sacan poco provecho de la desacentuación porque la interferencia es independiente de la frecuencia y no existe una ventaja *selectiva*. Para estar seguros, la desacentuación de alta frecuencia debe acarrear alguna mejoría, pero sólo lo suficientemente severa para justificar la complejidad consiguiente.

Volviendo a la FM en sí, el filtro de desacentuación es generalmente una red RC sencilla que tiene

$$H_{de}(f) = \left[1 + j\left(\frac{f}{B_{de}}\right)\right]^{-1} \approx \begin{cases} 1 & |f| \ll B_{de} \\ \dfrac{B_{de}}{jf} & |f| \gg B_{de} \end{cases} \tag{9}$$

† El sistema *Dolby* para grabación de cinta magnética va aún más lejos por medio del ajuste en forma dinámica de la cantidad de preacentuación/desacentuación en proporción inversa al contenido de alta frecuencia; véase Horowitz (1972).

FIGURA 7.6
Amplitud de interferencia detectada para FM con filtraje de desacentuación.

donde el ancho de banda a 3 dB $B_{de} = ½ \pi RC$ es mucho menor que el ancho de banda del mensaje W. Puesto que la amplitud de la interferencia aumenta en forma lineal con $|f_I|$ en ausencia de filtraje, la respuesta de interferencia desacentuada es $|H_{de}(f_I) \times f_I|$ como se bosqueja en la figura 7.6. Obsérvese que, como en la PM, esto resulta ser constante para $|f_I| \gg B_{de}$. Por lo que, la FM puede ser superior a la PM en cuanto a la interferencia, tanto de canal adyacente como la cocanal.

En el extremo transmisor, el correspondiente filtro de preacentuación es †

$$H_{pe}(f) = \left[1 + j\left(\frac{f}{B_{de}}\right)\right] \approx \begin{cases} 1 & |f| \ll B_{de} \\ \dfrac{jf}{B_{de}} & |f| \gg B_{de} \end{cases} \tag{10}$$

lo cual tiene poco efecto sobre las frecuencias bajas del mensaje. Sin embargo, a frecuencias mayores, el filtro actúa como un *diferenciador*, siendo el espectro de salida proporcional a $fX(f)$ para $|f| \gg B_{de}$. Pero diferenciar una señal antes de la modulación de frecuencia, ¡es equivalente a la *modulación de fase*! En consecuencia, la FM preacentuada es, en realidad, una combinación de FM y PM, ya que contiene las ventajas de ambas respecto a la interferencia. Como se podría esperar, esto vuelve a ser igualmente efectivo para reducir el *ruido*.

Refiriéndose a $H_{pe}(f)$ como se dio antes, se ve que la amplitud de la máxima frecuencia moduladora se incrementa en este mismo factor. En forma general, diremos que la desviación incrementada requiere mayor ancho de banda de transmisión, por lo que la mejoría por preacentuación-desacentuación no se logra sin pagar su precio. Por fortuna, muchas señales moduladoras de interés, en particular las señales de audio, tienen relativamente poca energía en el extremo de alta frecuencia de la banda del mensaje, y por ello las componentes de frecuencia más alta no desarrollan el máximo de desviación, determinándose el ancho de banda de transmisión por las componentes bajas de amplitud mayor. El agregar la precentuación de alta frecuencia tiende a igualar el espectro del mensaje, puesto que todas las componentes requieren el mismo ancho de banda. Bajo esta condición, el ancho de banda de transmisión no necesita incrementarse.

† En el problema 7.5 se da una realización de circuito.

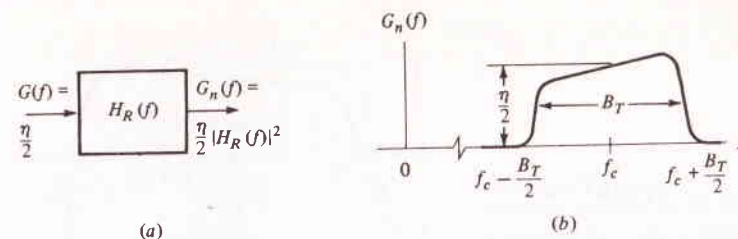



FIGURA 7.7
Ruido blanco filtrado pasabanda. (*a*) Diagrama a bloques; (*b*) densidad espectral de salida, $f \geq 0$.

EJERCICIO 7.2 Tómese $H_{pe}(f)$ en lugar de la ecuación (10) y justifíquese la afirmación de que la preacentuación no incrementa a B_T siempre y cuando la densidad espectral de $x(t)$ satisfaga

$$G_x(f) \leq \left(\frac{B_{de}}{f}\right)^2 G_{máx} \quad \text{para} \quad |f| > B_{de} \tag{11}$$

donde $G_{máx}$ es el valor máximo de $G_x(f)$ en $|f| < B_{de}$. ¿El espectro de voz típico de la figura 5.6*b*, satisface la ecuación (11)?.

7.3 RUIDO PASABANDA

La figura 7.7*a* aísla aquella parte de la figura 7.1 apropiada para un análisis del *ruido pasabanda* (o *de banda angosta*) $n(t)$ a la salida del filtro de predetección. Tomando $G(f) = \eta/2$ a la entrada, la densidad espectral de $n(t)$ es

$$G_n(f) = \frac{\eta}{2}\,|H_R(f)|^2 \tag{1}$$

como se bosqueja en la figura 7.7*b*, suponiendo a $H_R(f_c) = 1$ y $B_R \approx B_T$. Entonces, si el ancho de banda equivalente de ruido de $H_R(f)$ es aproximadamente igual a B_T, la potencia total del ruido filtrado es

$$N_R \triangleq \overline{n^2} = \sigma_n^2 = \eta B_T \tag{2}$$

y en lo sucesivo se le designará como potencia de ruido *recibida*. (Si suponemos que se trata de ruido blanco, la potencia de ruido queda indefinida antes del filtro de predetección).

Se presume, en vista de la naturaleza pasabanda de $n(t)$, que se le puede representar en las formas habituales

$$n(t) = R_n(t) \cos [\omega_c t + \phi_n(t)] \tag{3a}$$

$$= n_i(t) \cos \omega_c t - n_q(t) \operatorname{sen} \omega_c t \tag{3b}$$

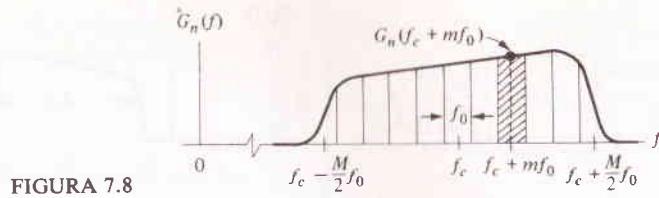



FIGURA 7.8

Pero $n(t)$ es una señal *aleatoria*, por lo que $R_n(t)$, $\phi_n(t)$, $n_i(t)$ y $n_q(t)$ deben ser también aleatorias. En consecuencia, el interés debe concentrarse en los promedios estadísticos de estas variadas y en sus funciones de densidad de probabilidad. Aunque no existe un camino directo de los espectros de potencia a las funciones de densidad de probabilidad, las técnicas avanzadas conducen a respuestas para este caso particular. Se dará una descripción cualitativa de la aproximación y se discutirán los resultados.

La sección 3.5 incluyó una representación en serie de Fourier de una señal aleatoria obtenida por la fragmentación de su espectro de potencia en bandas angostas. En concordancia, se fragmenta a $G_n(f)$ en $M = B_T/f_0$ bandas de ancho f_0 (Fig. 7.8) y recurriendo a la ecuación (16), (Sec. 3.5), se obtiene

$$n(t) \approx \sum_{m=-M/2}^{M/2} A_m \cos [(\omega_c + m\omega_0)t + \theta_m] \qquad (4)$$

donde

$$\overline{A_m{}^2} = 4f_0\, G_n(f_c + mf_0) = 2\eta f_0 |H_R(f_c + mf_0)|^2$$

y θ_m está uniformemente distribuida. Pero

$$\cos [(\omega_c + m\omega_0)t + \theta_m] = \cos (m\omega_0 t + \theta_m) \cos \omega_c t - \operatorname{sen}(m\omega_0 t + \theta_m) \operatorname{sen} \omega_c t$$

por lo que la ecuación (4) tiene la misma forma que la (3*b*) con

$$n_i(t) \approx \sum_{m=-M/2}^{M/2} A_m \cos (m\omega_0 t + \theta_m) \qquad (5a)$$

$$n_q(t) \approx \sum_{m=-M/2}^{M/2} A_m \operatorname{sen}(m\omega_0 t + \theta_m) \qquad (5b)$$

las cuales son las componentes en cuadratura del ruido pasabanda.

Componentes en cuadratura

Examinando las ecuaciones (5*a*) y (5*b*), se encuentra que el término de más alta frecuencia en cada una está en $Mf_0/2 = B_T/2$. Por lo general, $B_T/2$ es mucho menor que f_c y en consecuencia se puede decir que $n_i(t)$ y $n_q(t)$ son señales aleatorias *pasabajas* cuyas variaciones en el tiempo son lentas en comparación con la frecuencia portadora. Además, aunque las componentes en cuadratura no existen como entidades físicas en este punto, se debe asignar una densidad espectral de potencia a cada una.

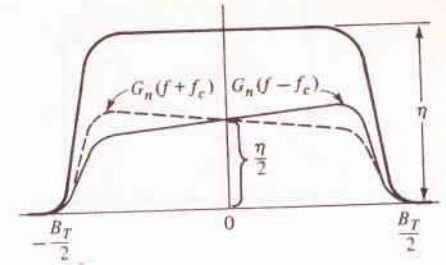



FIGURA 7.9
Densidad espectral de las componentes en cuadratura del ruido pasabanda, $G_{n_i}(f)$ y $G_{n_q}(f)$.

Para conseguirlo, primero se encuentra la función de autocorrelación de $n_i(t)$ aplicando la superposición a la ecuación (14), (Sec. 3.5), o sea,

$$R_{n_i}(\tau) \approx \sum_{m=0}^{M/2} 2f_0[G_n(f_c - mf_0) \cos(-m\omega_0 \tau) + G_n(f_c + mf_0) \cos m\omega_0 \tau]$$

$$\approx 2 \sum_{mf_0=0}^{B_T/2} [G_n(f_c - mf_0) + G_n(f_c + mf_0)] \cos (m\omega_0 \tau)\, f_0$$

Entonces, esta aproximación resulta ser exacta en el límite conforme $f_0 \to df$, $mf_0 \to$ F, y la suma conduce a la integración. Así

$$R_{n_i}(\tau) = 2 \int_0^{B_T/2} [G_n(f_c - f) + G_n(f_c + f)] \cos \omega\tau\, df \qquad (6$$

y $R_{n_q}(\tau)$ será lo mismo, puesto que no aparecen aquí los términos de fase θ_m. Si se pone $\tau = 0$, notando que $G_n(f_c \pm f) \approx 0$ para $f > B_T/2$, se desprende en forma rápida que

$$\overline{n_i^2} = \overline{n_q^2} = N_R \qquad (7$$

un resultado que no es inconsistente con la ecuación (2) dado que $\overline{n^2} = (\overline{n_i^2} + \overline{n_q^2})/2$

Recuérdese ahora el teorema de Wiener-Kinchine $R(\tau) \longleftrightarrow G(f)$ y el hecho de que $G(f)$ tiene simetría par cuando la señal en cuestión es real. En consecuencia se puede escribir

$$R(\tau) = \mathscr{F}^{-1}[F(f)] = 2 \int_0^{\infty} G(-f) \cos \omega\tau\, df$$

luego, comparando ésta con la ecuación (6), se tiene

$$G_{n_i}(f) = G_{n_q}(f) = G_n(f - f_c) + G_n(f + f_c) \qquad |f| < B_T/2 \qquad ($$

como se bosquejó en la figura 7.9. Muy a menudo, $H_R(f)$ ^{2} tiene simetría en relación con f_c y la ecuación (8) se simplifica a $G_{n_i}(f) = 2G_n(f - f_c)$, etc. En cualquier caso, la figura 7.9 refuerza el comentario reciente de que $n_i(t)$ y $n_q(t)$ son señales pasabajas.

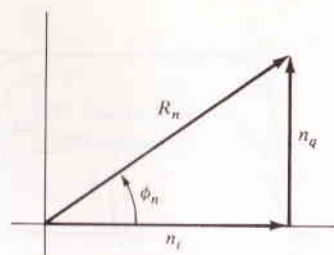



FIGURA 7.10
Diagrama fasorial para las componentes del ruido pasabanda.

Una inspección más a fondo de las ecuaciones (*5a*) y (*5b*) sugiere que es aplicable el *teorema del límite central*; n_i y n_q están constituidas por la suma de un número grande de variadas, cada una de las cuales proporciona una pequeña parte a la suma. La cuestión crucial es la dependencia estadística de los términos individuales. Un análisis más complejo revela que ellas son estadísticamente independientes. En consecuencias, n_i y n_q son de *distribución gaussiana* con valor medio cero y variancia N_R.

También se puede demostrar que n_i y n_q son ellas mismas estadísticamente independientes, conduciendo a la función de densidad de probabilidad conjunta

$$p(n_i,n_q) = p_{n_i}(n_i)p_{n_q}(n_q) = \frac{1}{2\pi N_R} e^{-(n_i+n_q)^2/2N_R} \tag{9}$$

Aunque $n_i(t)$ y $n_q(t)$ son estadísticamente independientes, ambas están relacionadas con $n(t)$. De manera específica,†

$$n_i(t) = n(t) \cos \omega_c t + \hat{n}(t) \operatorname{sen} \omega_c t$$
$$n_q(t) = \hat{n}(t) \cos \omega_c t - n(t) \operatorname{sen} \omega_c t$$

donde $\hat{n}(t)$ es la transformada de Hilbert de $n(t)$. Desde el punto de vista físico, las componentes en cuadratura (y las componentes de envolvente y fase) deben depender de y sólo de $n(t)$. Sin embargo, gracias a la independencia estadística, se puede tratar a n_i y a n_q como entidades distintas, siempre que se desee hacer más expedito un análisis.

EJERCICIO 7.3 Dedúzcase $\overline{n^2} = (\overline{n_i^2} + \overline{n_q^2})/2$ tomando el promedio del tiempo de ambos miembros de la ecuación (*3b*).

Envolvente y fase

Volviendo a la representación de envolvente y fase, por medio del diagrama fasorial común (Fig. 7.10), se relacionan las ecuaciones (*3a*) y (*3b*). Obviamente,

$$R_n^2 = n_i^2 + n_q^2 \qquad \phi_n = \arctan \frac{n_i}{n_q} \tag{10a}$$

† En Schwartz, Bennett y Stein (1966, Cap. 1) se trata en forma breve el ruido pasabanda, con mayores referencias.

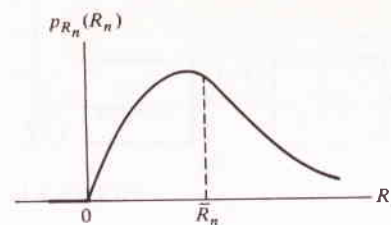



FIGURA 7.11
Función de densidad de probabilidad de Rayleigh.

y en forma recíproca

$$n_i = R_n \cos \phi_n \qquad n_q = R_n \operatorname{sen} \phi_n \tag{10b}$$

También recuérdese (Sec. 5.1), que las descripciones espectrales del caso de envolvente y fase, son difíciles. No obstante, es posible encontrar las funciones de densidad de probabilidad para R_n y ϕ_n por medio de la transformación funcional de la ecuación (9).

Puesto que las transformaciones funcionales que nos ocupan son no lineales se debe partir desde el principio escribiendo el equivalente bidimensional de la ecuación (11), (Sec. 3.2).

$$p_{n_i}(n_i)p_{n_q}(n_q)\,|dn_i\, dn_q| = p(R_n,\phi_n)\,|dR_n\, d\phi_n|$$

donde $p(R_n, \phi_n)$ es la función de densidad de probabilidad de R_n y ϕ_n. Insertando la ecuación (9) con $n_i^2 + n_q^2 = R_n^2$ y haciendo hincapié en que $dn_i dn_q = R_n dR_n d\phi_n$ se llega a

$$p(R_n,\phi_n) = \frac{|R_n|}{2\pi N_R} e^{-R_n^2/2N_R} \tag{11}$$

resultado que precisa dos requisitos: primero, por definición, R_n no puede ser negativa; segundo, se deben evitar las ambigüedades de fase limitando el intervalo de ϕ_n, digamos a $[-\pi, \pi]$. Así, $p(R_n, \phi_n) = 0$ para $R_n < 0$ o $|\phi_n| > \pi$.

La ecuación (11) es aún molesta, dado que ϕ_n no aparece en forma explícita. Sin embargo, de la ecuación (17), (Sec. 3.2) integrando con respecto a R_n se tiene

$$p_{\phi_n}(\phi_n) = \int_0^\infty p(R_n,\phi_n)\, dR_n = \frac{1}{2\pi} \qquad |\phi_n| < \pi \tag{12}$$

por lo que ϕ_n está *uniformemente distribuida* sobre $[-\pi, \pi]$, una conclusión más lógica de acuerdo a la figura 7.10 y las funciones de densidad de probabilidad de n_i y de n_q. Procediendo de manera semejante para la función de densidad de probabilidad de R_n, se integra con respecto a ϕ_n y se obtiene

$$p_{R_n}(R_n) = \frac{R_n}{N_R} e^{-R_n^2/2N_R} \qquad R_n \geq 0 \tag{13}$$

lo cual se conoce como la *distribución de Rayleigh*, bosquejada en la figura 7.11.

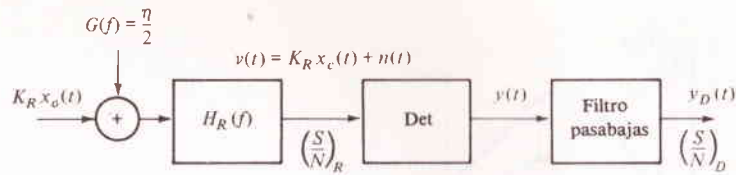



FIGURA 7.12

Obsérvese con sumo cuidado que N_R es la variancia de n (y n_i y n_q) pero no de R_n. En efecto

$$\overline{R_n^2} = 2N_R \qquad \overline{R_n} = \sqrt{\frac{\pi N_R}{2}} \tag{14}$$

mientras que

$$\overline{\phi_n^2} = \frac{\pi^2}{3} \qquad \overline{\phi_n} = 0 \tag{15}$$

como puede comprobar el lector.

Por último, de las ecuaciones (11) (12) y (13), $p(R_n, \phi_n) = p_{R_n}(R_n)p_{\phi_n}(\phi)$, lo cual significa que R_n y ϕ_n son *estadísticamente independientes*. Por lo que, se ha descrito al ruido pasabanda como una sinusoide aleatoria de frecuencia f_c cuya envolvente tiene una distribución de Rayleigh y cuya fase tiene una distribución uniforme. Las variaciones de envolventes y fase son estadísticamente independientes, sacándose en conclusión a partir de las ecuaciones (8) y (10) que ellas cambian en forma lenta en comparación con f_c.

EJERCICIO 7.4 Un aspecto afortunado de la distribución de Rayleigh es su facilidad de integración para la determinación de probabilidades. De manera específica, demuéstrese que

$$\text{Prob}\,[R_n > r] = e^{-r^2/2N_R} \tag{16}$$

[*Sugerencia*: Introdúzcase $\lambda = R_n^2/2N_R$ en $\int_r^{\infty} p(R_n)\,dR_n$.)

7.4 RUIDO EN LA MODULACIÓN LINEAL

Analizaremos ahora el comportamiento de los sistemas de modulación lineal ante el ruido. En la figura 7.12 se aprecia el modelo para el receptor, estando la señal modulada más el ruido pasabanda en la entrada del detector

$$v(t) = K_R x_c(t) + n(t) \tag{1}$$

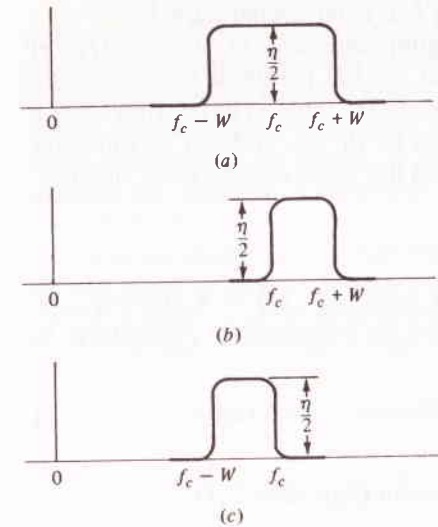



FIGURA 7.13
Espectro de ruido de predetección $G_n(f)$ en la modula lineal. (*a*) Doble banda lateral (*b*) banda lateral supe (*c*) banda lateral inferior.

lo cual es también válido para la modulación exponencial en $x_c(t)$ apropiada. Pue to que la modulación lineal tiene $B_T = 2W$ o W, dependiendo de que haya sido o n suprimida una banda lateral, los tres espectros de ruido posibles $G_n(f)$ son como lo que se muestran en la figura 7.13 considerando que $H_R(f)$ es simétrica.

La potencia promedio de la señal en la entrada del detector es $K_R^2 < x_c^2(t) >$ S_R, y $\overline{n^2} = N_R = \eta B_T$ es la potencia de ruido suponiendo que el ancho de banc equivalente de ruido del filtro de predetección es igual a B_T. La señal y el ruido so aditivos en la ecuación (1), por lo que resulta significativo definir a la ***relación señ a ruido de la predetección***

$$\left(\frac{S}{N}\right)_R \triangleq \frac{S_R}{N_R} = \frac{S_R}{\eta B_T}$$

lo cual lo sugiere el parámetro $\gamma = S_R/\eta W$ introducido en la sección 4.3; de manc específica, para valores iguales de S_R, η, y W,

$$\left(\frac{S}{N}\right)_R = \frac{W}{B_T}\gamma$$

En consecuencia, $(S/N)_R = \gamma$ para banda lateral única ($B_T = W$) mientras q $(S/N)_R = \gamma/2$ para doble banda lateral. Sin embargo, la interpretación que deb mos recordar es que γ es igual al valor máximo de la relación señal a ruido en el *de tino* $(S/N)_D$ en transmisión analógica de banda base. Por el mismo tenor, l ecuaciones (2) y (3) son en realidad *límites superiores*, o sea, $(S/N)_R < S_R/\eta B_T$ si $B > B_T$, etc. Por último, también se necesita

$$S_R = \frac{S_T}{\mathscr{L}} = K_R^2 S_T = \left(\frac{A_R}{A_c}\right)^2 S_T$$

lo cual relaciona a S_R con las expresiones dadas con anterioridad para S_T.

Con los preliminares ya dispuestos, la pregunta que surge es: dados $x_c(t)$ y el tipo de detector, ¿cuál es la señal final de salida, más la forma de onda de ruido $y_D(t)$?. Y si la señal y el ruido son aditivos en la salida, ¿cuál es la relación señal a ruido en el destino $(S/N)_D$, digamos en términos de $(S/N)_R$ o de γ? Se intentará dar contestación a estas preguntas investigando los dos casos de detección, sincrónica y de envolvente.

Detección sincrónica

Si la modulación es de doble banda lateral (DSB), entonces $x_c(t) = A_c x(t) \cos \omega_c t$ y $B_T = 2W$. Cuando se escribe al ruido en la forma de portadora en cuadratura, la ecuación (1) resulta ser

$$v(t) = \underbrace{[A_R x(t) + n_i(t)]}_{v_i(t)} \cos \omega_c t - n_q(t) \operatorname{sen} \omega_c t \tag{5}$$

así, recurriendo al modelo de detector de la ecuación (5*a*), (Sec. 7.1),

$$y(t) = v_i(t) = A_R x(t) + n_i(t)$$

En esta forma, después del filtraje pasabajas,

$$y_D(t) = A_R x(t) + n_i(t) \tag{6}$$

puesto que ambas componentes son de banda limitada en $W = B_T/2$.

Se pueden obtener tres conclusiones importantes de la ecuación (6). Primera, el mensaje y el ruido son aditivos en la salida; segunda, a la componente de ruido en cuadratura $n_q(t)$ se le rechaza por completo; tercera, el espectro de potencia de ruido en la salida es $G_{n_i}(f)$ la cual tiene la forma de $G_n(f)$ trasladada a la frecuencia cero. Si $|H_R(f)|$ es relativamente plana sobre B_T, el ruido de salida es en esencia blanco sobre el ancho de banda W del mensaje.

Tomando los valores cuadráticos medios de los dos términos de la ecuación (6), la relación señal a ruido en el destino es

$$\left(\frac{S}{N}\right)_D = \frac{A_R{}^2\overline{x^2}}{\overline{n_i{}^2}} = \frac{2S_R}{\eta B_T} = 2\left(\frac{S}{N}\right)_R \tag{7a}$$

puesto que $\overline{n_i{}^2} = \overline{n^2} = \eta - B_T$ y $A_R{}^2\overline{x^2} = 2S_R$ para DSB. De manera alterna, de la ecuación (3),

$$\left(\frac{S}{N}\right)_D = \frac{S_R}{\eta W} = \gamma \qquad \text{DSB} \tag{7b}$$

Por lo que, en cuanto al ruido se refiere, la DSB con detección sincrónica ideal es † equivalente a la transmisión en banda base, aún cuando el ancho de banda de transmisión sea dos veces mayor.

† Véase el problema 7.13 para los efectos de errores de fase en el sincronismo.

Desde el punto de vista de una traslación de frecuencia se podría haber sospechado un resultado diferente, a saber, $(S/N)_D = (S/N)_R = \gamma/2$, puesto que el espectro de la señal modulada y el espectro de ruido están desplazados hacia abajo, a la frecuencia cero. Sin embargo, las bandas laterales de la señal trasladada se traslapan en forma *coherente*, mientras que las bandas laterales del ruido se suman en forma *incoherente*. En efecto, es la coherencia de banda lateral la que equilibra la potencia de ruido de predetección duplicada $\eta B_T = 2\eta W$ en comparación con la banda base.

Los resultados para la AM demodulada en forma sincrónica se pueden inferir con prontitud de las ecuaciones (6) y (7), recordando que $x_c(t) = A_c[1 + mx(t)] \cos \omega_c t$ y la porción de la potencia de transmisión que conduce la información es $2p_{SB} = S_T - p_c = m\overline{x^2}\, A_c^2/2$. Por lo que

$$y_D(t) = A_R mx(t) + n_i(t) \tag{8}$$

y

$$\left(\frac{S}{N}\right)_D = \frac{A_R{}^2 m^2\overline{x^2}}{\eta B_T} = \frac{2m^2\overline{x^2}}{1 + m^2\overline{x^2}}\left(\frac{S}{N}\right)_R$$

$$= \frac{m^2\overline{x^2}}{1 + m^2\overline{x^2}}\gamma \qquad \text{AM} \tag{9}$$

Esta relación está limitada por $(S/N)_D \leq \gamma/2$, reflejando la observación hecha con anterioridad en el sentido de que el 50% o más de la potencia transmitida se desperdicia en la portadora de AM. Siendo iguales otros factores, un sistema de AM debe transmitir por lo menos dos veces más potencia que un sistema con portadora suprimida para conseguir la misma salida. Así, sobre la base de la potencia promedio, la AM resulta inferior a la DSB en 3 dB o más.

Bajo las mejores condiciones $(S/N)_D = \gamma/2$, correspondiendo a $m^2\overline{x^2} = 1$; con modulación de tono a carga máxima $m^2\overline{x^2} = \frac{1}{2}$ y $(S/N)_D = \gamma/3$. En forma más típica, sin embargo, $m^2\overline{x^2} \approx 0.1$, para lo cual $(S/N)_D$ está unos 7 dB abajo del valor máximo y 10 dB abajo de la DSB. La reducción consecuente de la relación señal a ruido es un serio problema en radiodifusión de AM. En consecuencia, muy seguido, se emplean en el transmisor técnicas especiales como compresión de volumen y limitación de picos para asegurar que la portadora sea modulada por completo la mayor parte del tiempo. Estas técnicas en realidad distorsionan al mensaje al ser recuperado y no serían aceptables para transmisión analógica de datos; en cambio, cuando el material de programa es de audio, la distorsión es más tolerable.

Para el caso de la modulación de banda lateral única (o residual), el análisis de la modulación de doble banda lateral se debe modificar por medio de una componente en cuadratura en $x_c(t)$ y de la frecuencia portadora de desviación permanente en comparación con la frecuencia central f_0 de $H_R(f)$. Con estos cambios

$$v(t) = \frac{A_R}{2}[x(t) \cos \omega_c t \mp \hat{x}(t) \operatorname{sen} \omega_c t] + [n_i(t) \cos \omega_0 t - n_q(t) \operatorname{sen} \omega_0 t] \tag{10}$$

donde $\omega_0 = 2\pi(f_c \pm W/2)$. Dejando los detalles al lector, la salida demodulada es

$$y_D(t) = \frac{A_R}{2} x(t) + n_i(t) \cos \pi Wt \pm n_q(t) \operatorname{sen} \pi Wt \qquad (11)$$

y

$$\left(\frac{S}{N}\right)_D = \frac{A_R{}^2\overline{x^2}}{4\eta B_T} = \left(\frac{S}{N}\right)_R = \gamma \qquad \text{SSB} \qquad (12)$$

puesto que B_T y $S_R = A_R^2\overline{x^2}/4$. Obsérvese que la componente de la señal en cuadratura $x(t)$ no tiene $y_D(t)$, como debe ser, pero el término de ruido en cuadratura aparece ahora en la salida. Esto se debe a que la frecuencia central del espectro de ruido $G_n(f)$ difiere de la frecuencia portadora en $\pm\ W/2$. Los argumentos de la traslación de frecuencia sugieren entonces que el ruido demodulado debe ser ruido "pasabanda" centrado en $W/2$. Los dos últimos términos de $y_D(t)$ tienen precisamente esta interpretación, por lo que el espectro de ruido de salida es, en esencia, constante sobre W excepto para un posible vacío en CD.

La ecuación (12) es también válida para banda lateral residual con portadora suprimida, siempre y cuando el ancho de la banda residual sea pequeño en comparación con el ancho de banda del mensaje. Para el caso de banda lateral residual más portadora se demuestra con facilidad que

$$\left(\frac{S}{N}\right)_D \approx \frac{m^2\overline{x^2}}{1 + m^2\overline{x^2}} \gamma \qquad \text{VSB + C} \qquad (13)$$

lo cual procede de las aproximaciones $B_T \approx W$ y $S_T \approx (1 + m^2\overline{x^2})\, A_c^2/2$. Si el lector desea, puede examinar por qué este resultado es el mismo que para AM, como en la ecuación (9). Revisando las ecuaciones de la (6) a la (13), se pueden establecer las siguientes características generales de la modulación lineal detectada en forma sincrónica.

1 El mensaje y el ruido son aditivos en la salida si lo son en la entrada del detector.
2 Si el espectro de ruido de predetección es razonablemente plano sobre el ancho de banda de transmisión, el espectro de ruido en el destino es esencialmente constante en el ancho de banda del mensaje.
3 En cuanto a las relaciones señal a ruido de salida, la modulación de banda lateral suprimida (SSB y VSB) no presenta una ventaja particular sobre la modulación de doble banda lateral (AM y DSB). Esto es así en razón de la propiedad de coherencia de la doble banda lateral, la cual compensa lo reducido de la potencia de ruido de predetección de la banda lateral única.
4 Permitiendo el "desperdicio" de potencia en los sistemas de portadora no suprimida, todos los tipos de modulación lineal tienen el mismo comportamiento que la transmisión de banda base sobre la base de potencia transmitida promedio y densidad de ruido fija.

Hasta este punto, se han comparado sistemas suponiendo valores iguales de S_R, la potencia *promedio*. Sin embargo, es quizá más realista tomar potencias *pico* adecuadas, reflejando la limitación de potencia pico de los transmisores prácticos. Bajo esta condición, la relación señal a ruido en el destino para DSB es cuatro veces la de AM (6 dB mejor) puesto que las potencias pico son proporcionales a A_c^2 y $4A_c^2$, respectivamente. Los cálculos de potencia pico para banda lateral única con señales moduladoras representativas indican† que la SSB es 2 a 3 dB mejor que la DSB y de 8 a 9 dB mejor que la AM. Pero la SSB es *inferior* a la DSB si el mensaje tiene discontinuidades pronunciadas que den lugar a "cuernos" de envolvente.

EJERCICIO 7.5 Desarróllense todos los detalles entre las ecuaciones (10), (11) y (12). (*Sugerencia*: La parte correspondiente al ruido de la ecuación (11) es ruido pasabanda en $W/2$, por lo que su valor cuadrático medio es igual a $(\overline{n_i^2} + \overline{n_q^2})/2$.)

Detección de envolvente y efecto de umbral

A medida que se demodula normalmente la AM por medio de un detector de envolvente, es necesario ver en cuánto difiere de la detección sincrónica cuando hay ruido. En la entrada del detector se tiene, como antes,

$$v(t) = A_R[1 + mx(t)] \cos \omega_c t + [n_i(t) \cos \omega_c t - n_q(t) \operatorname{sen} \omega_c t] \qquad (14)$$

La construcción fasorial de la figura 7.14*a* muestra que las resultantes de envolvente y fase son

$$R_v(t) = \left(\{A_R[1 + mx(t)] + n_i(t)\}^2 + [n_q(t)]^2\right)^{1/2}$$
$$\phi_v(t) = \arctan \frac{n_q(t)}{A_R[1 + mx(t)] + n_i(t)} \qquad (15)$$

Es obvio que un análisis más a fondo necesita algunas simplificaciones; en este caso, supóngase que la señal es muy grande o muy pequeña.

Suponiendo que la señal domina, digamos $A_R^2 \gg \overline{n^2}$, entonces $A_R[1 + mx(t)]$ será grande comparada con $n_i(t)$ y $n_q(t)$, por lo menos la mayor parte del tiempo. La envolvente se puede aproximar por medio de

$$R_v(t) \approx A_R[1 + mx(t)] + n_i(t) \qquad (16)$$

la cual muestra la modulación por el ruido, similar a la de interferencia. La aproximación procede de la figura 7.14*a* o del desarrollo de la ecuación (15). En consecuencia

$$y_D(t) = R_v(t) - \overline{R}_v = A_R mx(t) + n_i(t)$$

† Downing (1964, Cap. 4).

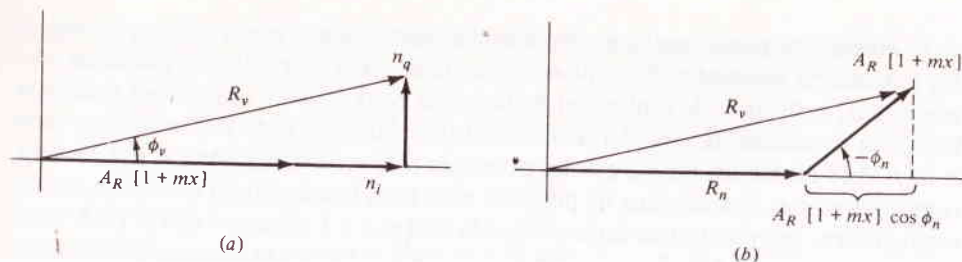



FIGURA 7.14
Diagramas fasoriales para AM más ruido. (*a*) $(S/N)_R \gg 1$; (*b*) $(S/N)_R \ll 1$.

idéntica a la expresión de la detección sincrónica, ecuación (8). La relación señal a ruido de la postdetección será entonces como se dio previamente en la ecuación (9). Se puede hacer un análisis similar para modulación de banda lateral de señal compatible, SSB + C o VSB + C, con resultados similares.

Pero se debe recordar la condición para estos resultados, a saber, $A_R^2 \gg \overline{n^2}$. Puesto que $A_R^2/\overline{n^2}$ es proporcional a $S_R/\eta B_T$, un requisito equivalente es $(S/N)_D \gg 1$. (No existe tal condición en la detección sincrónica.) Así, siempre y cuando la relación señal a ruido de la predetección sea grande, la demodulación de envolvente en presencia del ruido tiene el mismo tipo de comportamiento que la demodulación sincrónica.

En el otro extremo, o sea cuando $(S/N)_R \ll 1$, la situación es bastante diferente. Porque si $A_R^2 \ll \overline{n^2}$, el ruido domina de una manera similar al caso de fuerte interferencia, y se puede pensar que $x_c(t)$ modula a $n(t)$ en vez de a la inversa. Para facilitar el análisis, $n(t)$ se representa en la forma de envolvente y fase

$$n(t) = R_n(t)\cos[\omega_c t + \phi_n(t)]$$

conduciendo al diagrama fasorial de la figura 7.14*b*. En esta figura el fasor ruido es la referencia, pues se está considerando a $n(t)$ como dominante. Por inspección, la envolvente es en forma aproximada

$$R_v(t) \approx R_n(t) + A_R[1 + mx(t)]\cos\phi_n(t) \tag{17}$$

de la cual

$$y(t) = R_n(t) + A_R m x(t)\cos\phi_n(t) - \overline{R}_n \tag{18}$$

La componente principal de la salida es en forma obvia la envolvente de ruido $R_n(t)$, como se esperaba. Además, no existe en la ecuación (18) un término estrictamente proporcional al mensaje; aunque la señal y el ruido fueron *aditivos* en la entrada, el término del mensaje detectado se *multiplica* por el ruido en la forma de $\phi_n(t)$, el cual es aleatorio. El mensaje queda inesperadamente *mutilado*, y la información se pierde. En efecto, se puede decir que el mensaje *no existe* en la salida, puesto que no se puede recuperar. Bajo estas circunstancias, una relación señal a ruido en la salida resulta difícil de definir, si no es que sin sentido.

La mutilación o pérdida del mensaje a bajas relaciones señal a ruido de predetección, recibe el nombre de *efecto de umbral*. El nombre resulta apropiado porque existe un valor determinado de $(S/N)_R$ sobre el cual la mutilación es despreciable y debajo del cual es comportamiento del sistema se deteriora en forma rápida. En esta forma, si la señal transmitida se desvanece o se incrementa el ruido del receptor, el mensaje desaparece muy rápidamente. Obsérvese que el efecto de umbral ocurre sólo con la detección de envolvente. En la detección sincrónica, la señal de salida y el ruido son siempre aditivos, pues el mensaje queda sepultado en el ruido si $(S/N)_R \ll 1$, pero se conserva su identidad.

En realidad, el umbral no es un punto exclusivo a menos que se establezca cierta convención para definirlo. En términos generales, los efectos de umbral son mínimos si $A_R \gg R_n$ la mayor parte del tiempo. Para especificar más el caso se definirá el *nivel de umbral* como aquel valor de $(S/N)_R$ para el cual $A_R \geq R_n$ con 0.99 de probabilidad, es decir $[R_n > A_R] = 0.01$. De la ecuación (16), (Sec. 7.3),

$$\text{Prob}\,[R_n > A_R] = e^{-A_R^2/2N_R} \approx e^{-S_R/2N_R}$$

donde se ha puesto $m^2\overline{x^2} \approx 1$, por lo que $A_R^2 \approx S_R$. Resolviendo $e^{-S_R/2N_R} = 0.01$ para el nivel de umbral, se tiene

$$(S/N)_{R\text{-}\ll\, esima} \tag{19a}$$

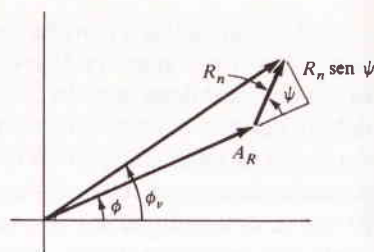
o, puesto que $(S/N)_R = \gamma/2$,

$$\text{gamma-ésima} \tag{19b}$$

Si $(S/N)_R < (S/N)_{R\text{-ésima}}$ (o $\gamma <$ *gamma*-ésima), se debe esperar mutilación del mensaje, con la pérdida de información consecuente.

Revisando el valor de $(S/N)_{R\text{-ésima}}$ y recordando que $(S/N)_D < (S/N)_R$ se llega a una conclusión significativa: el efecto de umbral casi nunca limita seriamente a los sitemas de AM. Aclaremos este concepto: una inteligibilidad razonable en la transmisión de voz necesita una relación señal a ruido de postdetección de alrededor de 30 dB o más, $(S/N)_D \geq 1\,000$, para lo cual $(S/N)_R$ es buena sobre el nivel de umbral. En otras palabras, el ruido aditivo oscurece a la señal mucho antes de que el ruido multiplicativo la mutile. Por otra parte, existen técnicas complejas de procesamiento para recuperar señales *digitales* sepultadas en ruido aditivo. En consecuencia, si se emplea la AM para transmisión digital, puede ser necesaria la detección sincrónica para evitar los efectos de umbral.

Por último, resulta importante considerar cómo puede actuar un detector de envolvente en forma sincrónica y por qué requiere una $(S/N)_R$ grande. Refiriéndonos a la figura 5.30, y suponiendo que el ruido de entrada es despreciable, se ve que las funciones del diodo son las de un conmutador. Si cerramos brevemente durante los picos positivos de la portadora, la conmutación queda sincronizada perfectamente con la portadora, pero cuando predomina el ruido, la conmutación es controlada sobre todo por los picos de ruido y por ello se pierde el sincronismo. Este último efecto nunca ocurre en los detectores sincrónicos verdaderos, donde la portadora generada en forma local puede siempre ser mucho mayor que el ruido.

FIGURA 7.15
Diagrama fasorial para FM o PM más ruido, $(S/N)_R \gg 1$.

EJERCICIO 7.6 Una señal de voz con $\overline{x^2} = 0.2$ y $W = 3$ kHz va a ser transmitida empleando la modulación lineal en un sistema que tiene $S_T = 600$ W, $\mathscr{L} = 30$ dB, y $\eta = 10^{-8}$ W/Hz. Suponiendo que no existe limitación de ancho de banda, selecciónese un tipo de modulación apropiado y encuéntrese $(S/N)_D$ en decibeles. Repítase para $\mathscr{L} = 40$ dB.

7.5 RUIDO EN LA MODULACIÓN EXPONENCIAL

Volviendo a la demodulación de FM o PM contaminadas con ruido, la situación es la misma que en la figura 7.12 con $x_c(t) = A_c \cos [\omega_c t + \phi(t)]$ donde

$$\phi(t) = \phi_\Delta x(t) \qquad \text{PM}$$
$$\frac{1}{2\pi}\frac{d\phi(t)}{dt} = f_\Delta x(t) \qquad \text{FM} \tag{1}$$

En uno u otro caso, el ruido pasabanda es simétrico respecto de f_c con $N_R = \eta B_T$ mientras que $S_R = A_R^2/2$ por lo que

$$\left(\frac{S}{N}\right)_R = \frac{A_R^2}{2\eta B_T} \tag{2}$$

Con $n(t)$ escrita en la forma de envolvente y fase, la entrada del detector es

$$v(t) = A_R \cos [\omega_c t + \phi(t)] + R_n(t) \cos [\omega_c t + \phi_n(t)] \tag{3}$$

e inmediatamente se ven las dificultades analíticas para encontrar la fase resultante para ser insertada en el modelo matemático del detector. Hágase como se procedió con la detección de envolvente y supóngase que la componente de señal domina al ruido.

La figura 7.15 muestra el diagrama fasorial con la diferencia de fase

$$\psi(t) = \phi_n(t) - \phi(t)$$

Tomando la aproximación por ángulo pequeño usual

$$\phi_v(t) \approx \phi(t) + \frac{R_n(t)}{A_R} \operatorname{sen} \psi(t) \tag{4}$$

siempre y cuando $(S/N)_R \gg 1$. No debe sorprender el que el primer término de la ecuación (4) sea la modulación del mensaje, pues el segundo término contiene tanto al mensaje como al ruido y es otra fuente de dificultad que se debe tratar en el momento oportuno. Mientras tanto, aplicando las ecuaciones (5*c*) y (5*d*), (Sec. 7.1), a $\phi_v(t)$ se llega a la señal demodulada

$$y(t) = \begin{cases} \phi_\Delta x(t) + \xi_{\text{PM}}(t) & \text{PM} \\ f_\Delta x(t) + \xi_{\text{FM}}(t) & \text{FM} \end{cases} \tag{5}$$

donde se ha definido

$$\xi_{\text{PM}}(t) = \frac{R_n(t)}{A_R} \operatorname{sen} \psi(t) \tag{6a}$$

$$\xi_{\text{FM}}(t) = \frac{1}{2\pi}\frac{d\xi_{\text{PM}}(t)}{dt} \tag{6b}$$

Interpretando a éstos como los términos *de ruido de postdetección*, aunque ellos dependen de la señal moduladora, la ecuación (5) tiene la forma de señal más ruido, es decir, ruido *aditivo* en la salida. Más aún,

$$S_D = \begin{cases} \phi_\Delta^2 \overline{x^2} & \text{PM} \\ f_\Delta^2 \overline{x^2} & \text{FM} \end{cases} \tag{7}$$

es la potencia de la señal de salida.

A pesar de estas conclusiones, no se pueden escribir en seguida las relaciones señal a ruido $(S/N)_D$, porque la evaluación de $\overline{\xi^2}$ es una tarea más que trivial, porque no se han incluido aquí los efectos del filtraje de postdetección sobre el ruido. Por ambas razones, se debe ir al dominio de la frecuencia y examinar los espectros de potencia de ruido en la postdetección.

Características del ruido en la postdetección

Se obtiene una descripción cualitativa de los espectros de ruido en la postdetección recordando que se puede pensar del ruido que es una suma de sinusoides aleatorias distribuidas igualmente sobre $f_c \pm B_T/2$. Dado que se supuso que $(S/N)_R \gg 1$, las amplitudes de estas sinusoides son muy pequeñas y cada una actuará como una señal interferente pequeña. Recurriendo a los estudios sobre la interferencia y a la figura 7.4, se infiere que el espectro de potencia de ruido en la postdetección de PM será constante sobre $|f| < B_T/2$. Pero para FM, donde la amplitud de la interferencia se incrementa en forma lineal con la frecuencia, el espectro de potencia irá como frecuencia *al cuadrado*. Este *espectro parabólico de potencia* refleja la noción de que las componentes de ruido de la entrada cercanas a la frecuencia portadora no pueden provocar tanta desviación de frecuencia como aquellas que se encuentran más lejos de f_c.

Yendo de lo cualitativo a lo cuantitativo, se calculará $G_\xi(f)$ a partir de la ecuación (6) haciendo caso omiso de la modulación de portadora y sustituyendo a

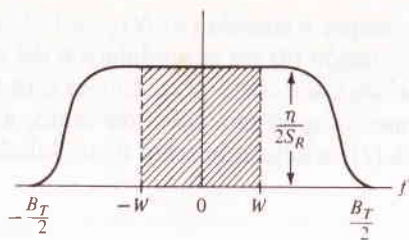



FIGURA 7.16
Espectro de ruido de postdetección en la PM, $G\xi_{PM}(f)$.

$\psi(t) = \phi_n(t) - \phi(t)$ por $\phi_n(t)$ sola. La aproximación se justifica dado que ϕ_n está uniformemente distribuida; en consecuencia, en el sentido de los promedios de ensamble, $\phi_n - \phi$ difiere de ϕ_n sólo en un corrimiento de valor medio. Con esta simplificación, $R_n(t)$ sen $\psi(t)$ se reduce a la componente de ruido en cuadratura $n_q(t) = R_n(t)$ sen $\phi_n(t)$, y ruido detectado viene a ser

$$\xi_{PM}(t) = \frac{1}{A_R} n_q(t) \tag{8a}$$

$$\xi_{FM}(t) = \frac{1}{2\pi A_R} \frac{dn_q(t)}{dt} = \frac{1}{2\pi} \frac{d\xi_{PM}(t)}{dt} \tag{8b}$$

El que éstas no dependan de la componente de ruido en fase $n_i(t)$ está de acuerdo con la interpretación fasorial de la modulación exponencial (Fig. 6.7), donde se dice que sólo las componentes en cuadratura con la portadora producen modulación de fase o de frecuencia.

Enfocando primero el caso de la PM, se encuentra, de la ecuación (8*a*), que la densidad espectral de ruido es $G_{\xi PM}(f) = G_{nq}(f)/A_R^2$. Pero $A_R^2 = 2S_R$ y, de las ecuaciones (1) y (8), (Sec. 7.3), $G_{nq}(f) = (\eta/2)[|H_R(f-f_c)|^2 + |H_R(f+f_c)|^2]$. Poniendo éstas juntas, se tiene que

$$G_{\xi_{PM}}(f) = \frac{\eta}{4S_R} [|H_R(f-f_c)|^2 + |H_R(f+f_c)|^2] \tag{9a}$$

$$\approx \frac{\eta}{2S_R} \Pi\left(\frac{f}{B_T}\right) \tag{9b}$$

donde la ecuación (9*b*) supone que $|H_R(f)| \approx 1$ sobre $f_c \pm B_T/2$, generalmente sería cierto si despresiamos al factor de ganancia que se cancelará en las relaciones señal a ruido. Bajo esta condición, $G_{\xi PM}(f)$ es de manera esencial plana sobre $|f| < B_T/2$ (Fig. 7.16). Pero $B_T/2$ excede normalmente el ancho de banda del mensaje, significando que las componentes de ruido fuera de banda—como

† Downing (1964, Cap. 5) demuestra que cuando $\phi(t)$ es de importancia, existen componentes adicionales en $G\xi(f)$ pero a frecuencias superiores de la banda del mensaje. Tales componentes se eliminan por medio del filtro de salida y no se deben tomar en cuenta.

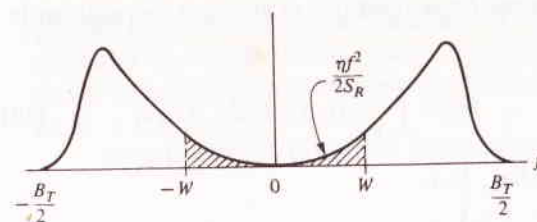



FIGURA 7.17
Espectro de ruido de postdetección en la FM, $G\xi_{FM}(f)$.

la interferencia fuera de banda— deben ser removidas por medio de filtraje de postdetección. Por consiguiente, con un filtro pasabajas ideal a la salida del detector, la potencia de ruido en el destino es

$$N_D = \int_{-W}^{W} G_{\xi PM}(f)\, df$$

$$= \int_{-W}^{W} \frac{\eta}{2S_R}\, df = \frac{\eta W}{S_R} \qquad \text{PM} \tag{10}$$

empleando la ecuación (9*b*) para $G_{\xi PM}(f)$.

Habiendo encontrado $G_{\xi PM}(f)$, la FM resulta ser un caso de rutina en vista de la ecuación (8*b*). De manera concreta, si aplicamos la relación de diferenciación para espectros de potencia, ecuación (27), (Sec. 2.6), se obtiene

$$G_{\xi_{FM}}(f) = \left(\frac{1}{2\pi}\right)^2 (2\pi f)^2 G_{\xi_{PM}}(f) \tag{11a}$$

$$\approx \frac{\eta f^2}{2S_R} \Pi\left(\frac{f}{B_T}\right) \tag{11b}$$

la cual, como se bosquejó en la figura 7.17, va como f^2 sobre la banda del mensaje. Esta figura, con la 5.37*b*, muestra por qué la FM estéreo adolece de más ruido que la FM monoaural, puesto que, antes de la modulación de subportadora, la señal diferencia trasladada $x_i(t) - x_R(t)$ reside entre los 23 y los 53 kHz, donde la densidad espectral del ruido es mucho mayor que en el intervalo de 0 a 15 kHz de la señal monoaural. De manera similar, cuando se diseña algún sistema de multicanalización por división en frecuencia con modulación de portadora en FM, la característica de *ruido de postdetección parabólica*, debe ser tomada en cuenta.

Dada la potencia total de ruido en la postdetección, se supone de nuevo que $|H_R(f)|$ es razonablemente plana, por lo que es válida la ecuación (11*b*). Por lo que, con un filtro pasabajas ideal se remueve el ruido fuera de banda en la salida,

$$N_D = \int_{-W}^{W} G_{\xi_{FM}}(f)\, df = \frac{\eta W^3}{3S_R} \qquad \text{FM} \tag{12}$$

sin filtraje de desacentuación.

Agregando un filtro de desacentuación se modifica el espectro de ruido en la salida y, en general,

$$N_D = \int_{-W}^{W} |H_{de}(f)|^2 G_{\xi_{FM}}(f)\, df \qquad (13)$$

Para el filtro RC de la ecuación (9), (Sec. 72.),

$$N_D = \int_{-W}^{W} \left[1 + \left(\frac{f}{B_{de}}\right)^2\right]^{-1} \left(\frac{\eta f^2}{2S_R}\right) df$$

$$= \left(\frac{\eta B_{de}^{3}}{S_R}\right)\left[\left(\frac{W}{B_{de}}\right) - \arctan\left(\frac{W}{B_{de}}\right)\right] \qquad (14a)$$

$$\approx \frac{\eta B_{de}^2 W}{S_R} \qquad \frac{W}{B_{de}} \gg 1 \qquad (14b)$$

donde la aproximación procede de que $(W/B_{de}) \approx \pi/2 \ll W/B_{de}$. Comparando la ecuación (14*b*) con la (12), se demuestra que la desacentuación ha reducido sustancialmente el ruido de salida, es decir,

$$\frac{N_D \text{ (con desacentuación)}}{N_D \text{ (sin desacentuación)}} \approx 3\left(\frac{B_{de}}{W}\right)^2 \qquad \frac{B_{de}}{W} \ll 1 \qquad (15)$$

Un análisis similar para la modulación de fase con desacentuación de alta frecuencia da un factor de reducción menos impresionante de aproximadamente $\pi B_{de}/2W$.

EJERCICIO 7.7 Encuéntrese el factor de reducción de ruido para modulación de fase con un filtro de desacentuación RC. *Resp.:* (B_{de}/W) arctan ½(W/B_{de})

Relaciones señal a ruido en el destino

Encontradas las potencias de la señal de salida y del ruido, al fin se llega a las relaciones señal a ruido en el destino y a una tasa de la *reducción del ruido de banda ancha* en los sistemas de modulación exponencial. Se supone que $(S/N)_R \gg 1$ en todo y que cualquier filtraje de desacentuación se compensa por medio de un filtraje de preacentuación pues se presume que el espectro del mensaje es tal, que esta última operación no incrementa en forma significativa el ancho de banda de transmisión.

Empezando con la modulación de fase, S_D y N_D se dan en las ecuaciones (7) y (10), respectivamente, por lo que

$$\left(\frac{S}{N}\right)_D = \frac{\phi_\Delta^2\overline{x^2}}{(\eta W/S_R)} = \phi_\Delta^2\overline{x^2}\frac{S_R}{\eta W} = \phi_\Delta^2\overline{x^2}\gamma \qquad \text{PM} \qquad (16)$$

Puesto que γ es la relación señal a ruido de la salida para transmisión en banda base (o modulación lineal de portadora suprimida) con potencia S_R, ancho de banda W, y densidad de ruido η, se desprende que la modulación de fase añade una mejoría respecto a la banda base exactamente de $\phi_\Delta^2\overline{x^2}$. Pero dada la limitación por ambigüedad $\phi_\Delta \le \pi$, la mejoría que añade la PM no es mayor que $\phi_\Delta^2\overline{x^2}$/máx $= \pi^2$, o alrededor de 10 dB cuando mucho. Por supuesto, si $\phi_\Delta^2\overline{x^2} < 2$, el comportamiento de la PM es inferior a la banda base, mientras que el ancho de banda de transmisión es $B_T \ge 2W$.

Para la FM sin filtraje de desacentuación, las ecuaciones (7) y (12) originan

$$\left(\frac{S}{N}\right)_D = \frac{f_\Delta^2\overline{x^2}}{(\eta W^3/3S_R)} = 3\left(\frac{f_\Delta}{W}\right)^2 \overline{x^2}\frac{S_R}{\eta W}$$

En consecuencia, insertando la relación de desviación $\Delta = f_\Delta/W$,

$$\left(\frac{S}{N}\right)_D = 3\Delta^2\overline{x^2}\gamma \qquad \text{FM} \qquad (17)$$

Si agregamos un filtro RC de desacentuación $B_{de} \ll W$ se reduce aún más el ruido de salida alrededor de $3(B_{de}/W)^2$, por lo que

$$\left(\frac{S}{N}\right)_D = \frac{3\Delta^2\overline{x^2}\gamma}{3(B_{de}/W)^2} = \left(\frac{W}{B_{de}}\right)^2 \Delta^2\overline{x^2}\gamma$$

$$= \left(\frac{f_\Delta}{B_{de}}\right)^2 \overline{x^2}\gamma \qquad B_{de} \ll W \qquad (18)$$

Pero con o sin desacentuación, la FM presenta una reducción sustancial del ruido de banda ancha tal, que la relación señal a ruido en el destino se incrementa con la relación de desviación y el grande ancho de banda de transmisión concomitante. En efecto, sucede que la relación señal a ruido en el destino se puede hacer *arbitrariamente grande* incrementando sólo la desviación; ésta es una conclusión que requiere un estudio más a fondo que se volverá a examinar adelante.

Para recalcar la dependencia del ancho de banda de transmisión, recuérdese que la FM de banda ancha ($\Delta \gg 1$) tiene $B_T = 2f_\Delta = 2\Delta W$, por lo que $\Delta^2 = B_T^2/4W^2$, y la ecuación (17) viene a ser

$$\left(\frac{S}{N}\right)_D = \frac{3}{4}\left(\frac{B_T}{W}\right)^2 \overline{x^2}\gamma \quad \text{FM de banda ancha} \qquad (19)$$

Bajo esta condición, el comportamiento del sistema mejora como el *cuadrado* de la relación de anchos de banda B_T/W. Con desviaciones más pequeñas el punto de ruptura par, comparado con banda base, ocurre en $2\Delta^2\overline{x^2}\,3 = 1$; en consecuencia, la

$$\Delta = (3\overline{x^2})^{-1/2} \approx 0.6$$

algunas veces se le designa como la línea divisoria entre las FM de banda angosta y de banda ancha.

El monto preciso de lo que se puede ganar con la FM de banda ancha, queda evidenciado con la radiodifusión de FM comercial, para la cual $f_\Delta = 75$ kHz, $W =$ 15 kHz $\Delta = 5$, el filtro de desacentuación normal tiene $B_{de} = 2.1$ kHz, correspondiendo a la constante de tiempo RC $= 75\ \mu$s.

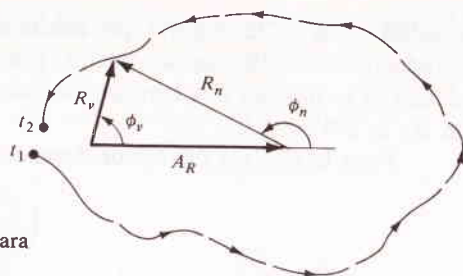



FIGURA 7.18
Diagrama fasorial y lugar geométrico para $A_R{}^2 \approx R_n{}^2$.

Tomando a $\overline{x^2} = ½$ como valor representativo y excluyendo la desacentuación, la relación señal a ruido en el destino es $(3 \times 5^2 \times ½)\gamma = 37.5\,\gamma$, o alrededor de 16 dB mejor que en banda base. Incluyendo la mejoría por desacentuación, se tiene $(S/N)_D = 640\,\gamma$ y, continuando iguales los otros factores, un sistema de FM de 1 W sustituiría a un sistema de banda base de 640 W sin reducción de la relación señal a ruido en la salida. Pero no se olvide que el ancho de banda de transmisión es 14 veces más grande que el ancho de banda del mensaje, subrayando la transacción de ancho de banda por potencia o relación señal a ruido que hace posible la modulación exponencial.

Si tomamos en cuenta este intercambio ancho de banda-potencia en FM, veremos que en la práctica existen varios factores que se oponen a la completa realización de una potencia de transmisión reducida a expensas de un ancho de banda de transmisión incrementado. Además, el objetivo de la FM comercial no es reducir la potencia transmitida, sino propiciar la mejor relación posible entre señal y ruido de salida. No obstante, existen muchas otras ocasiones en que el mínimo de potencia es esencial; para tales aplicaciones la condición $(S/N)_R \gg 1$ resulta ser un lastre, y el efecto de umbral en FM se presenta como un asunto de graves consecuencias.

Efecto de umbral en FM

Todos los resultados anteriores se basaron en la aproximación dada por la ecuación (4), en la que se supone una relación señal a ruido grande en la predetección. Cuando, por otra parte, $(S/N)_R \ll 1$, la fase resultante en la entrada del detector es

$$\phi_v(t) \approx \phi_n(t) + \frac{A_R}{R_n(t)} \operatorname{sen}[\phi(t) - \phi_n(t)] \qquad (20)$$

obviamente el mensaje —contenido en $\phi(t)$— ha sido mutilado por el ruido más allá de toda esperanza de recuperarlo.

En realidad, la mutilación empieza a ocurrir cuando $(S/N)_R \approx 1$ porque entonces $A_R{}^2 \approx R_n{}^2$, por lo que los fasores de señal y de ruido son casi iguales en longitud, y las variaciones de la fase del ruido ocasionan variaciones comparables de $\phi_v(t)$. Un caso de particular interés se muestra en el diagrama fasorial de la figura 7.18, en que se omite la modulación $\phi(t)$, donde $\phi_v(t_1) \approx \phi_n(t_1) \approx -\pi$ mientras que $\phi_v(t_2) \approx +\pi$. Si la variación de fase entre t_1 y t_2 sigue el lugar geométrico delineado,

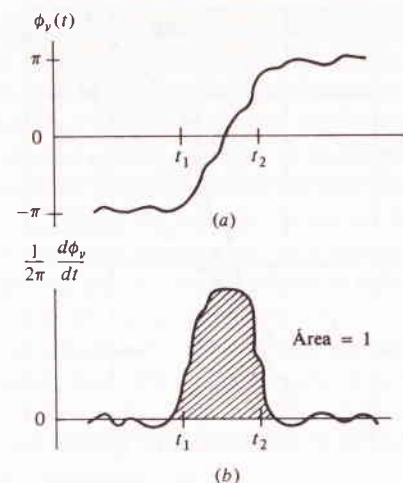



FIGURA 7.19
Variaciones de ruido típicas cerca del umbral. (*a*) Desviación de fase con escalón; (*b*) desviación de frecuencia con pico.

entonces $\phi_v(t)$ se ve como en la figura 7.19*a* y la correspondiente desviación de frecuencia $f_v(t) = (1/2\pi)\, d\phi_v(t)/dt$ es la figura 7.19*b*, un *espigón* de área unitaria y duración mínima $(t_2 - t_1) \times$ mín $\approx 2/B_T$, (¿por qué?). En forma auditiva, estos picos o espigones ocasionan crujidos o chasquidos que desvirtúan la señal deseada.

De esta gráfica cuantitativa se infiere que el espectro de ruido de salida no es muy parabólico, pero tiende a llenarse de CD, y los picos de salida tienen un apreciable contenido de baja frecuencia. Esta conclusión ha sido verificada por medio de análisis detallados y para ello se empleó la aproximación de "chasquido" perfeccionada por Rice †. El análisis es complicado (y fuera de este enfoque) porque las características del pico cambian cuando la portadora se modula, o sea, el efecto conocido también como de supresión de modulación. Así, se obtienen resultados cuantitativos sólo para señales moduladoras específicas. En el caso de la modulación de tono, el ruido de salida total resulta ser

$$N_D = \frac{\eta W^3}{3S_R}\left[1 + \frac{12\Delta}{\pi}\gamma e^{-(W/B_T)\gamma}\right] \qquad (21)$$

donde el segundo término es la contribución de los picos.

La figura 7.20 muestra a $(S/N)_D$ en decibeles y su gráfica contra γ (también en decibeles) para dos valores de la relación de desviación Δ, empleando modulación de tono y N_D dada por la ecuación (21). El abatimiento súbito de estas curvas, trazadas al factor exponencial de la ecuación (12), es el efecto de umbral de la FM. Cuando el sistema está operando cerca de la "rodilla" de una curva, pequeñas variaciones de potencia de la señal causan cambios alternos de tamaño mediano en la señal de

† Véanse Rice (1948) y Stumpers (1948) para el trabajo original o el texto de Taub y Schilling (1971, Cap. 10).

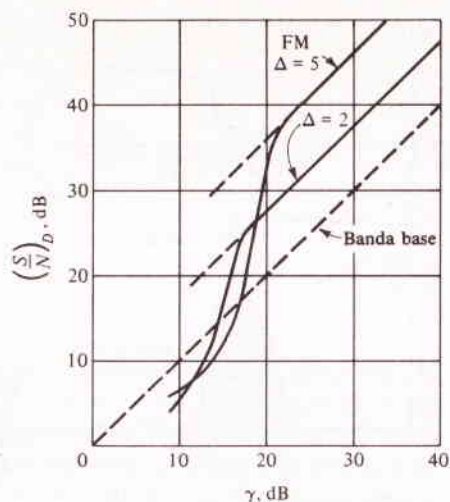



FIGURA 7.20
El comportamiento del ruido en FM como una función de $\lambda = S_R/\eta W$, sin incluirse la desacentuación.

salida. Esta pérdida súbita en la salida no es común en la zona marginal de la recepción de FM †.

Empero de la figura 7.20 es, en cierta forma engañosa, pues implica una señal de salida utilizable debajo de la curva en forma de rodilla, mientras que una señal moduladora arbitraria sufriría mutilación. Estudios experimentales indican que la mutilación resulta despreciable en muchos casos si $(W/B_T)\gamma = (S/N)_R \geq 10$ tiene este valor o uno semejante. En consecuencia, se define que el punto de umbral está en

$$(S/N)_{R\text{-ésima}} = 10 \tag{22}$$

por lo que

$$\gamma_{\text{ésima}} = 10\,\frac{B_T}{W} = 20M(\Delta) \tag{23a}$$

$$\approx 20(\Delta + 2) \qquad \Delta > 2 \tag{23b}$$

donde se ha hecho uso de la ecuación del ancho de banda de FM, $B_T = 2M(\Delta)\,W \approx 2(\Delta + 2)\,W$. Las ecuaciones (22) y (23) se aplican también a la modulación fase, donde Δ reemplaza a ϕ_Δ.

La figura 7.20 demuestra en forma correcta que el comportamiento de la FM sobre el umbral es bastante impreciso para llamar la atención, pues la transmisión en banda base en el mejor de los casos da $(S/N)_D = \gamma$. La figura 7.20 no incluye la mejoría adicional que se logra por medio del filtraje de desacentuación, pero obsérvese lo que sucede si se intenta hacer a $(S/N)_D$ arbitrariamente grande si incremen-

† Ocurre un fenómeno similar con la interferencia co-canal de dos señales que tienen amplitudes casi iguales en el receptor. Pequeñas variaciones de una amplitud relativa originan que la más intensa domine en forma súbita la situación, desplazando a la otra completamente en la salida. A esto se le ha dado el nombre apropiado de *efecto de captura*.

tamos sólo la relación de desviación y mantenemos fija γ, digamos en 20 dB. Con $\Delta = 2(B_T \approx 7W)$ se llega poco arriba del umbral, y $(S/N)_D \approx 28$ dB; pero con $\Delta = 5(B_T \approx 14W)$ se está bajo el umbral, y la señal de salida no es útil a causa de la mutilación. Por lo tanto, no se puede lograr un trueque ilimitado de ancho de banda por relación señal a ruido, y si se incrementa la desviación puede resultar un serio deterioro en el comportamiento del sistema.

Hacer un trueque de ancho de banda por potencia reducida, resulta también prohibitivo. Supóngase, por ejemplo, que se desea una relación señal a ruido de 30 dB con un mínimo de potencia transmitida y que el ancho de banda de transmisión puede ser tan grande como $B_T = 14W$. De no ser por el umbral, se podría emplear FM con $\Delta = 5$ y $\gamma = 14$ dB, un ahorro de potencia de 16 dB en comparación con la banda base. Pero el punto de umbral para $\Delta = 5$ es γ-ésima $= 21.5$ dB, para el cual $(S/N)_D = 37.5$ dB. Así, lo que dicta el diseño es el punto de umbral más que la relación señal a ruido deseada, y la virtual reducción de potencia no se puede realizar por completo.

En vista de estas consideraciones, es útil calcular $(S/N)_D$ en el punto de umbral. Así, omitiendo de nuevo la desacentuación, se sustituye la ecuación (23) en la (17) para obtener

$$\left(\frac{S}{N}\right)_{D\,\text{ésima}} = 3\Delta^2\overline{x^2}\,\gamma_{\text{ésima}} \tag{24a}$$

$$\approx 60\Delta^2(\Delta + 2)\overline{x^2} \qquad \Delta > 2 \tag{24b}$$

el cual es el valor mínimo de $(S/N)_D$ como una función de Δ. Dado un valor específico para $(S/N)_D$ y sin limitación de ancho de banda, se puede resolver la ecuación (24) para la relación de desviación Δ que da lugar al *desempeño más eficiente* en términos de la potencia de la señal. Por supuesto que se debe hacer alguna concesión en cuanto a un posible desvanecimiento de la señal, ya que no es aconsejable operar sin un margen relativo para el punto de umbral.

EJERCICIO 7.8 Diséñese un sistema de FM de potencia mínima sin limitación de ancho de banda tal que $(S/N)_D = 50$ dB, dados $x^2 = 1/2$, $W = 10$ kHz y $\eta = 10^{-8}$ W/Hz. *Resp.*: $\Delta \approx 14$, $B_T \approx 320$ kHz, $S_R \geq 32$ mW.

Extensión de umbral en FM (FMFB) ★

Como la limitación de umbral constituye un serio inconveniente en el diseño de sistemas de FM de potencia mínima, se ha suscitado un considerable interés por las técnicas de *extensión de umbral*, sobre todo para la comunicación espacial, pero mucho antes de que aparecieran estas últimas aplicaciones, Chaffee (1939) propuso un medio para extender el punto de umbral en FM empleando un lazo de realimentación de seguimiento de frecuencia o de compresión de frecuencia en el receptor, es decir, realimentación en FM (FMFB). Aquí se bosqueja la idea básica de este sistema. En

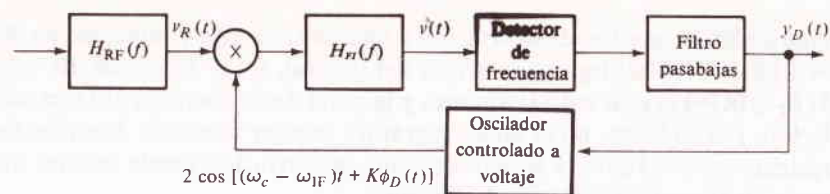



FIGURA 7.21
El receptor de FM con realimentación (FMFB)

a literatura† se pueden encontrar otras técnicas de extensión de umbral —en forma notable, el receptor con lazo de sujeción de fase (PLL)— con un análisis detallado de la FMFB.

Un receptor de FMFB es un superheterodino en el cual se sustituye al oscilador local por un oscilador controlado a voltaje cuya frecuencia instantánea de salida se controla por medio de la señal demodulada, dando la configuración de realimentación de la figura 7.21. En la forma cuantitativa, el oscilador sigue o rastrea la frecuencia instantánea de la señal modulada de tal manera que, como se podrá ver, la desviación efectiva de frecuencia en la entrada del amplificador de FI se reduce y el ancho de banda de la FI puede ser mucho menor que B_T (pero, por supuesto, no menor que $2W$). Si $B_{FI} \ll B_T$, la potencia del ruido en la entrada del detector es mucho menor que N_R y, en consecuencia, se reduce el nivel de umbral.

De manera específica, $N_{FI} = (B_{FI}/B_T)N_R$ y se evita la mutilación si $S_R/N_{FI} = (B_T/B_{FI})(S/N)_R \geq 10$. Por lo tanto, el nuevo nivel de umbral es

$$\left(\frac{S}{N}\right)_{R\text{-ésima}} = 10\,\frac{B_{FI}}{B_T}$$

$$\gamma\text{-ésima} = 10\,\frac{B_{FI}}{B_T}\,\frac{B_T}{W} = 10\,\frac{B_{FI}}{W} \geq 20 \qquad (25)$$

en comparación con *gamma*-ésima $= 20M(\Delta)$ para un receptor convencional. Falta demostrar que B_{FI} puede ser menor que B_T y que el comportamiento arriba del umbral no ha sufrido degradación. Esto se hará primero en el caso de una señal modulada, sin ruido, y después el de una portadora no modulada, con ruido.

En la figura 7.21, la salida del oscilador controlado a voltaje es igual a $2\cos[(\omega_c - \omega_{FI})t + K\phi_D(t)]$ donde $\omega_{FI} = 2\pi f_{FI}$, K es una constante, y

$$\phi_D(t) = 2\pi\int_{-\infty}^{t} y_D(\lambda)\,d\lambda$$

† Taub y Schilling (1971, Cap. 10); Schwartz, Bennett y Stein (1966, Cap. 3).

En ausencia de ruido, la señal recibida $v_R(t)$ es

$$v_R(t) = A_R\cos[\omega_c t + \phi(t)] \qquad \text{con} \quad \phi(t) = 2\pi f_\Delta\int_{-\infty}^{t} x(\lambda)\,d\lambda$$

Por lo que, suponiendo que $H_{FI}(f)$ remueve el término suma de frecuencias de la salida del mezclador, la entrada al detector será la señal diferencia de frecuencias

$$v(t) = A_R\cos[\omega_{FI}t + \underbrace{\phi(t) - K\phi_D(t)}_{\phi_v(t)}]$$

Así

$$y_D(t) = \frac{1}{2\pi}\frac{d\phi_v(t)}{dt} = f_\Delta x(t) - Ky_D(t)$$

y resolviendo para $y_D(t)$ se tiene

$$y_D(t) = \frac{f_\Delta}{1+K}x(t) \qquad (26)$$

por lo que la desviación de frecuencia se ha reducido $1/(1+K)$, por el efecto de seguimiento de la frecuencia de la realimentación. Esta reducción de la desviación también se tiene en la salida del mezclador, puesto que $\phi_v(t) = 2\pi[f_\Delta/(1+K)]\phi(t)$; en conformidad, el ancho de banda del amplificador de FI puede ser

$$B_{FI} = 2M(\Delta_{FI}) \qquad \Delta_{FI} = \frac{f_\Delta}{(1+K)W}$$

y $B_{FI} \approx 2W$ si se elige a K tal que $B_{FI} \ll 1$.

Ahora considérese una portadora no modulada, más ruido con $S_R/N_{FI} \gg 1$. Si $n_{qFI}(t)$ es el ruido en cuadratura a la salida del amplificador de FI, resulta que la densidad espectral de éste es constante sobre B_{FI} si $B_{FI} \ll B_T$ por lo que

$$G_{n_{qFI}}(f) = \eta\,|H_{FI}(f - f_{FI})|^2$$

Entonces, de las ecuaciones (25), (8*b*) y (12) se infiere que el ruido demodulado es

$$\xi(t) = \frac{1}{(1+K)2\pi A_R}\frac{dn_{qFI}(t)}{dt}$$

y

$$N_D = \frac{\eta W^3}{(1+K)^2 3S_R}$$

Por último, de la ecuación (26), $S_D = f_\Delta^2\overline{x^2}/(1+K)^2$ por lo que $(S/N)_D$ es la misma que para un receptor convencional, ecuación (17), puesto que $(1+K)^2$ se cancela en la relación.

Para no extraviar al estudiante, algunos factores han sido tratados en forma superficial en este análisis general. En realidad, se ha apreciado el comportamiento en las cercanías del umbral, y se ha omitido el tema completo sobre el rastreo de errores y transitorios. La conclusión de la relación señal a ruido en el destino arriba del umbral es correcta y en consecuencia de la reducción del ancho de banda. Sin embargo, resulta que el punto de umbral depende de la relación de desviación Δ que no se incluye en la ecuación (25). Por eso, la extensión de umbral de FMFB queda limitada a aproximadamente 5 a 7 dB factor apreciable, aun en diseños de potencia mínima.

7.6 COMPARACIÓN DE SISTEMAS DE MODULACIÓN EN ONDA CONTINUA (CW)

Al fin llegamos a una posición desde la cual podemos comparar bien los tipos de modulación de onda continua. La tabla 7.1 resume los puntos comparables: ancho de banda de transmisión normalizado, $\mathscr{B} \triangleq B_T/W$; relación señal a ruido en el destino $(S/N)_D$ normalizada por γ; punto de umbral, si hay alguno; respuesta de CD o de baja frecuencia; y complejidad de instrumentación. La tabla incluye también la transmisión en banda base para propósitos de referencia. Como en lo anterior, $\gamma \triangleq S_R/\eta W$ donde W es el ancho de banda del mensaje, S_R es la potencia de la señal recibida y $\eta = k\mathscr{T}_N$ es la densidad de ruido referida a la entrada del receptor. Se suponen sistemas muy cercanos al ideal, por lo que los valores de $(S/N)_D$ son límites superiores.

De los diferentes tipos de modulación lineal, los métodos de portadora suprimida son superiores a la AM convencional en varios aspectos; las relaciones señal a ruido son mejores y no existe efecto de umbral. Cuando es importante la *conservación del ancho de banda*, son particularmente recomendables la banda lateral única y la banda lateral residual. Pero raras veces se consigue algo por nada en este mundo, y el precio de una modulación lineal eficiente es la gran complejidad de instrumentación, de manera especial en el receptor. La detección sincrónica, sin importar cómo se logre, necesita circuitos muy complejos en comparación con el detector de envolvente. Para comunicación *punto a punto* (un transmisor, un receptor) el precio puede ser digno de tomarse en cuenta. Pero para sistemas de *radiodifusión* (un transmisor, *muchos* receptores) las consideraciones económicas inclinan la balanza hacia el receptor más simple posible, y por ende, la detección de envolvente.

Desde un punto de vista de la instrumentación, la AM es la modulación lineal menos compleja, mientras que la VSB con portadora suprimida, con su filtro de banda lateral especial y requisitos de sincronización, es más compleja. Respecto de la DSB y la SSB (en sus aplicaciones apropiadas) lo último es menos difícil de instrumentar, dado que la sincronización no es tan crítica. Además, los avances de la tecnología han hecho posible que los filtros de banda lateral se consigan con mayor facilidad. Asimismo, la VSB + C está clasificada de complejidad "moderada", a despecho del filtro vestigial, puesto que la detección de envolvente es suficiente.

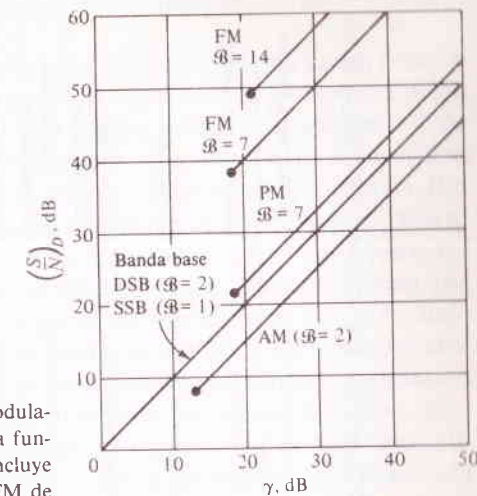



FIGURA 7.22
Comportamiento de los sistemas de modulación de CW (onda continua) como una función de $\gamma = S_R/\eta W \gamma$ $\mathscr{B} = B_T/W$; se incluye una mejora por desacentuación para FM de 12 dB.

Comparada con la banda base o la modulación lineal, la modulación exponencial puede proporcionar incrementos sustanciales en los valores de $(S/N)_D$ —sobre todo en la FM con desacentuación— con sólo una instrumentación moderadamente compleja. La figura 7.22 ilustra esto en forma similar a la 7.20 (tomando de nuevo $\overline{x^2} = ½$) excepto que se ha agregado a las curvas de FM una mejora típica por desacentuación de 12 dB y se ha omitido el comportamiento debajo del umbral. Todas las curvas están rotuladas con la relación de anchos de banda $\mathscr{B}$.

Es obvio que para valores iguales de $\mathscr{B}$ la FM es marcadamente superior a la PM en cuanto al desempeño contra el ruido. (La modulación de fase, desde luego, tiene sus ventajas sobre la FM para la transmisión de señales *digitales*, lo que será analizado en el Cap. 10.) A medida que el sistema se extienda por encima del umbral, la mejoría se podrá aumentar paulatinamente incrementando a $\mathscr{B}$, mientras que la PM está limitada a $\mathscr{B} \leq 10$, puesto que $\phi_\Delta \leq \pi$.

El precio a pagar por esta mejoría es un ancho de banda de transmisión excesivo. Por eso es más apropiada la modulación exponencial de banda ancha cuando se desean señales de salida extremadamente limpias y la conservación del ancho de banda es un factor secundario. Las aplicaciones típicas incluyen a la FM comercial y audio en televisión. En frecuencias de microondas, resultan ventajosas las propiedades tanto de reducción de ruido como de amplitud constante, por lo que muchos sistemas de microondas tienen modulación de portadora exponencial.

En cuanto a la conservación de potencia, la FM con valores moderados de $\mathscr{B}$ ofrece un ahorro sobre la modulación lineal, a pesar de las limitaciones de umbral. En este carácter se emplea para comunicación de radio, punto a punto, especialmente en sistemas móviles, donde la detección sincrónica sería una molestia. Más aún, gracias a las técnicas de extensión de umbral, la FM de banda ancha se emplea en

Tabla 7.1 COMPARACIÓN DE LOS SISTEMAS DE MODULACIÓN EN ONDA CONTINUA (CW).

Tipo	$\mathscr{B} = B_T/W$	$(S/N)_D/\gamma$	γ-ésimo	CD	Complejidad	Comentarios	Aplicaciones típicas
Banda base	1	1	…	No[a]	Menor	Sin modulación	Enlaces de corta distancia
AM	2	$\frac{m^2\overline{x^2}}{1+m^2\overline{x^2}}$	20	No	Menor	Detección de envolvente $m \leq 1$	Radiodifusión
DSB	2	1	…	Sí	Mayor	Detección sincrónica	Datos analógicos, multicanalización
SSB	1	1	…	No	Moderada	Detección sincrónica	Voz de punto a punto, multicanalización
VSB	1+	1	…	Sí	Mayor	Detección sincrónica	Datos digitales
VSB + C	1+	$\frac{m^2\overline{x^2}}{1+m^2\overline{x^2}}$	20	Sí[b]	Moderada	Detección de envolvente $m < 1$	Video de televisión
PM	$2M(\phi_\Delta)$	$\phi_\Delta^2\overline{x^2}$	$10\mathscr{B}$	Sí	Moderada	Detección de fase $\phi_\Delta \leq \pi$	Datos digitales
FM[d]	$2M(\Delta)$	$3\Delta^2\overline{x^2}$	$10\mathscr{B}$	Sí	Moderada	Detección de frecuencia	Radiodifusión, enlaces de microondas con repetidores

[a] A menos que sea de acoplamiento directo.
[b] Con restauración electrónica de CD.
[c] $\mathscr{B} \geq 2$.
[d] Sin incluir la desacentuación.

ciertos sistemas de comunicación por satélite y sondas espaciales, donde la potencia del transmisor es invaluable.

Después de observar la transmisión de señales moduladoras que tienen componentes significativas de baja frecuencia, se probó la superioridad de la modulación de doble banda lateral y de lateral residual; esto explica su uso en la transmisión de datos, tanto analógica como digital. Para facsímil y video de TV, la restauración electrónica de la CD hace posible y deseable la VSB con detección de envolvente. (La AM también se podría emplear en este caso, pero el ancho de banda es prohibitivo. La banda lateral con portadora suprimida queda, virtualmente, fuera de controversia.) También, en la sección 6.5 se dijo que un discriminador balanceado tiene una excelente respuesta de baja frecuencia; por lo que el comportamiento a baja frecuencia de la FM se puede considerar igual al de la DSB o de la VSB, y sin la problemática sincronización. En consecuencia, la FM de desviación pequeña o la modulación de subportadora de PM se emplean a menudo en sistemas de telemetría por las señales de entrada menos activas. Por razones similares, las grabadoras de cinta magnética de alta calidad están equipadas con un modo de FM en el cual se graba la entrada como una onda modulada en frecuencia.

En la tabla no se muestra el comportamiento relativo del sistema en relación con las características de transmisión variando en el tiempo, el desvanecimiento de frecuencia selectiva, la propagación multitrayectoria, etc. Un medio de transmisión inestable tiene un efecto *multiplicativo*, el cual es particularmente desastroso para la detección de envolvente. (Los radioescuchas nocturnos que gustan sintonizar estaciones de AM distantes están familiarizados con los resultados alterados por mutilación en las audiciones.) De manera similar, las inestabilidades en la transmisión a menudo impiden la modulación de banda ancha.

Resumir en pocas palabras es imposible. No existe una solución universal para todos los problemas de comunicación. Por lo tanto, el ingeniero en comunicaciones debe acercarse a cada nuevo problema con la mente abierta y hacer una revisión cuidadosa de toda la información de que se dispone.

7.7 PROBLEMAS

7.1 (Secc. 7.1) Demuéstrese que las ecuaciones de la (5*a*) a la (5*d*) son modelos apropiados al tomar $v(t) = K_R x_c(t)$ donde $x_c(t)$ tiene el tipo de modulación correspondiente, es decir, DSB o SSB, AM, PM y FM, respectivamente.

7.2 (Secc. 7.2) Encuéntrese una expresión aproximada para la envolvente de una onda de AM de tono modulado más una sinusoide interferente $A_i \cos[(\omega_c + \omega_i)t + \theta]$ con $A_i \ll A_R$.

7.3 (Secc. 7.2) Investíguese el comportamiento de una detección de envolvente de AM contra la detección sincrónica de AM en presencia de la propagación multitrayectoria tal que $v(t) = x_c(t) + \alpha x_c(t - t_d)$, donde $\alpha^2 < 1$ y $x_c(t) = [1 + mx(t)]\cos\omega_c t$. Subráyense las diferencias cuando $\omega_c t_d \approx n\pi$ y $\omega_c t_d \approx (n + \frac{1}{2})\pi$.

7.4 (Secc. 7.2) Empezando con la ecuación (1), demuéstrese que la desviación instantánea

de frecuencia de $v(t)$ es $f_v(t) = m_l f_l(m_l + \cos \omega_l t)(1 + m_l t)(1 + m_l^2 + 2m_l \cos \omega_l t)$ donde $m_l = A_l/A_R$. Tomando a $m_l = 0.5$, bosquéjese $f_v(t)$ para $0 \le \omega_l t \le 180°$.

7.5 (Secc. 7.2) La figura P7.1 es el *circuito de preacentuación* normal cuando se emplea para desacentuación un filtro pasabajas RC. Encuéntrese $H_{pe}(f)$ y demuéstrese que se aproxima a la ecuación (10) dentro de una constante de proporcionalidad cuando $2\pi RC = 1/B_{de}$ y $R_0 \ll R$.

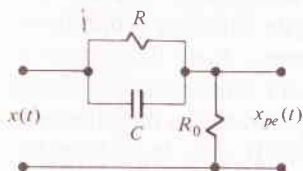



FIGURA P7.1

7.6★ (Secc. 7.2) Cuando no se satisface la ecuación (11), el filtraje de preacentuación incrementará el ancho de banda de la señal de FM. Se puede estimar el incremento empleando la *desviación de frecuencia rms* $f_{\Delta_{pe}} = f_\Delta(\overline{x_{pe}^2})^{1/2}$, donde $x_{pe}(t)$ es la salida de $H_{pe}(f)$ y se supone que $x(t)$ y $x_{pe}(t)$ son ambas de banda limitada en W con $\bar{x} = \bar{x}_{pe} = 0$ y $\overline{x^2} = 1$. El ancho de banda con preacentuación se calcula después empleando $f_{\Delta_{pe}}$ en lugar de f_Δ. (*a*) Tomando $H_{pe}(f) = K/H_{de}(f)$ y $H_{de}(f)$ por la ecuación (9), demuéstrese que $(f_{\Delta_{pe}}/f_\Delta)^2 = K^2[1 + \int_{-\infty}^{\infty}(f/B_{de})^2 G_x(f)df]$.
(*b*) Evalúese $B_{T_{pe}}/B_T$ cuando $K = 1$, $G_x(f) = (½\ W)\Pi(f/2W)$, $B_{de} = W/3$ y $f_\Delta = 5W$.

7.7 (Secc. 7.3) Bosquéjese $G_{n_i}(f)$ cuando $\eta = 40$ y $H_R(f)$ como se ha construido la gráfica en la figura P5.1.

7.8 (Secc. 7.3) Supóngase que $H_R(f)$ tiene simetría con $\pm f_c$ y $B_N \approx B_T \ll f_c$, por lo que se puede escribir $|H_R(f)^2| = |H_{LP}(f - f_c)^2| + |H_{LP}(f + f_c)|^2$ donde $H_{LP}(f) = H_R(f + f_c)\, u(f + f_c)$. Demuéstrese que la autocorrelación de $n(t)$ es $R_n(\tau) = \eta B_T \varrho(\tau) \cos 2\pi f_c \tau$ donde $(\tau) = (1/B_T)\mathscr{F}^{-1}[H_{LP}(f)^2]$, y justifíquese la aproximación $(\tau) \approx 1$ para $|\tau| \ll 1/B_T$.

7.9 (Secc. 7.3) Desarróllense todos los detalles entre las ecuaciones (9) y (11).

7.10 (Secc. 7.3) Considérese la señal aleatoria $y(t) = 2n(t)\cos(\omega_c t + \theta)$ donde $n(t)$ es ruido pasabanda centrado en f_c. Analícense y discútanse las propiedades de $y(t)$, dando atención a su dependencia de θ.

7.11 (Secc. 7.3) Se procesa ruido pasabanda por medio de un dispositivo cuya salida $y(t)$ es el cuadrado de la envolvente de entrada. Encuéntrense $p_y(y)$, $\bar{y}$ y $\overline{y^2}$.
Respuesta: $p_y(y) = (2\sigma_n^2)^{-1} e^{-y/2\sigma_n^2} u(y)$.

7.12 (Secc. 7.3) A través de un limitador riguroso pasa, ruido pasabanda con $B_T \ll f_c$. Bosquéjese una muestra típica de la salida $y(t)$ y escríbase una expresión para $p_y(y)$.

7.13 (Secc. 7.4) El demodulador de producto de la figura 5.27 tiene DSB más ruido en la entrada y existe un *error de fase* ϕ' en el oscilador por lo que su salida es $A_{OL} \cos(\omega_c t + \phi')$. Encuéntrese $y_D(t)$ y demuéstrese que $(S/N)_D$ está reducida en el factor $\cos^2\phi'$.

7.14 (Secc. 7.4) Un receptor de banda lateral superior tiene $|H_R(f)|^2$ como se muestra en la figura P7.2. Bosquéjese el espectro de ruido de salida y encuéntrese $(S/N)_D$ en términos de γ.

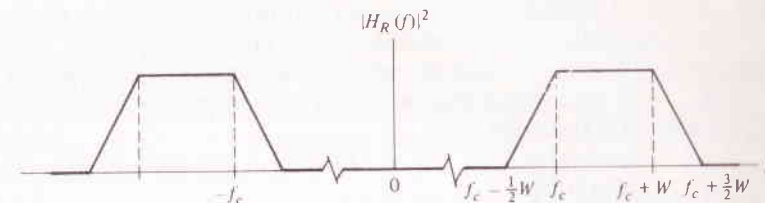



FIGURA P7.2

7.15★ (Secc. 7.4) Analícese el comportamiento de un *detector homodino* de DSB (Fig. 5.29), cuando la salida tiene ruido aditivo pasabanda. Tómese $x_c(t) = [x(t) + a] \cos\omega_c t$, donde $a \cos\omega_c t$ es la portadora piloto, y sea el filtro piloto uno pasabanda centrado en f_c con ancho de banda $B_p \ll 2W$.

7.16 (Secc. 7.4) Un sistema de AM con detección de envolvente tiene una $(S/N)_D = 30$ dB bajo condiciones de modulación de tono a plena carga, con $W = 8$ kHz. Si todos los anchos de banda se incrementan en forma conveniente mientras los otros parámetros permanecen fijos, ¿cuál es el mayor valor de W para el cual el sistema queda sobre el umbral?

7.17 (Secc. 7.4) Un receptor *heterodino* es un superheterodino sin filtraje de RF. Con ayuda de la figura 5.33, explíquese por qué $(S/N)_D$ puede estar a 3 dB abajo que un receptor superheterodino comparable. Examínense las diferencias cuando la mayor parte del ruido se genera dentro del receptor en comparación con el que entra con la señal.

7.18 (Secc. 7.4) Empléense consideraciones de ruido para explicar los siguientes enunciados. (*a*) En un receptor de banda lateral única es deseable tener a $|H_R(f)|$ rectangular en forma aproximada con $B_R = B_T$, mientras que esto no es crítico para la DSB.
(*b*) En un receptor de AM con detección de envolvente es deseable tener $|H_R(f)|$ muy cercana a la rectangular con $B_R = B_T$, mientras que esto no es crítico para detección sincrónica de AM.

7.19 (Secc. 7.4) Sea $v(t) = 4\cos\omega_c t + n(t)$, donde $n(t)$ es ruido pasabanda centrado en $f_c + 20$ con $B_T = 8$ y $\eta = 0.02$. Obténgase una expresión aproximada para $R_v(t)$ y evalúese R_v^2.

7.20 (Secc. 7.5) Bosquéjese $G_{\xi FM}(f)$ cuando $|H_R(f)|$ es como en el problema 6.25 con $f_0 = f_c$ y $Q = f_c/B_T$.

7.21 (Secc. 7.5) Las ecuaciones (10) y (12) son para un filtro pasabajas ideal. Encuéntrese N_D para FM y PM cuando el filtro de postdetección es un filtro pasabajas Butterworth de orden M-ésimo con $|H(f)| = [1 + (f/W)^{2M}]^{-1/2}$. Supóngase que $|H_R(f - f_c)|^2 \approx 1$ para $|f| \le B_T/2$ y $B_T \gg 2W$.

7.22 (Secc. 7.5) Estímese el factor de reducción de ruido para FM con un filtro gaussiano de desacentuación $H_{de}(f)^2 = \exp[-(\ln 2)(f/B_{de})^2$ donde $B_{de} \ll W \ll B_T/2$. Compárese con la ecuación (15).

7.23 (Secc. 7.5) Cierto sistema de modulación de fase tiene una $(S/N)_D = 30$ dB. Si la modulación se cambia a FM con el mismo B_T y se agrega un filtro de desacentuación RC con $B_{de} = W/10$, ¿cuál es el nuevo valor de $(S/N)_D$?
Respuesta: 50 dB.

7.24 (Secc. 7.5) La señal $x(t) = \cos 2\pi 200t$ se envía en FM sin desacentuación. Existe un filtro pasabanda ideal que pasa $100 \le |f| \le 300$ en la salida del discriminador. Calcúlese $(S/N)_D$ si $f_\Delta = 1$ kHz y $S_R/\eta = 500$.

7.25 (Secc. 7.5) Modifíquese la ecuación (24) para incluir filtraje de desacentuación RC y repítase el ejercicio 7.8 tomando $B_{de} = 2$ kHz.

7.26★ (Secc. 7.5) Con referencia al problema 7.6*a*, supóngase que K se elige de tal manera que $f_{\Delta_{pe}} = f_\Delta$, por lo que la desacentuación no incrementa en forma significativa el ancho de banda de transmisión.
(*a*) Demuéstrese que $S_D = (Kf_\Delta)^2\overline{x^2}$ y por ello $(S/N)_D = f_\Delta{}^2\overline{x^2}\gamma/[B_{de}{}^2 + \int_{-\infty}^{\infty} f^2 G_x(f)\,df]$.
(*b*) Simplifíquese esta expresión suponiendo que $G_x(f) = [f_0/\pi(f^2 + f_0{}^2)]\Pi(f/2W)$ donde $f_0 = B_{de} \ll W$. Compárese con las ecuaciones (17) y (18).

7.27 (Secc. 7.5) Los picos o chasquidos en la FM cerca del umbral, también ocurren cuando hay interferencia con $A_I \approx A_R$. Demuéstrese esto bosquejando a $\dot{f}v(t)$ en forma burda en el problema 7.4 con $m_I = 1 + \epsilon$, $\epsilon^2 \ll 1$. (Téngase cuidado en las cercanías de $\cos \omega_I t = \sigma^{-1}$.) ¿Qué sucede cuando $\epsilon \to 0$?

7.28 (Secc. 7.6) Un sistema de comunicación tiene $\mathscr{L} = 100$ dB, $\eta = 10^{-14}$ W/Hz, $\overline{x^2} = 1$, y $W = 10$ kHz. Calcúlese el valor de S_T en kilowatts que se requieren para $(S/N)_D \geq 40$ dB cuando la modulación es:
(*a*) SSB
(*b*) AM con $m = 1$ y $m = 0.1$
(*c*) PM con $\phi_\Delta = \pi$
(*d*) FM con $f_\Delta = 10$, 50 y 100 kHz
Omítase la desacentuación en el caso de la FM, pero compruébese para las limitaciones de umbral.
Respuesta: (*a*) 10; (*b*) 20, 1, 010; (*c*) 1; (*d*) 3.3, 0.14, 0.24.

7.29 (Secc. 7.6) Repítase el problema 7.28 con $W = 20$ kHz.

7.30 (Secc. 7.6) Una señal con $\overline{x^2} = ½$ se transmite en AM con $m = 1$ y una $(S/N)_D = 13$ dB. Si la modulación se cambia a FM (sin desacentuación) y se incrementan los anchos de banda mientras los otros parámetros permanecen fijos, encuéntrese el mayor valor útil de Δ y la $(S/N)_D$ resultante.

7.31★ (Secc. 7.6) Un sistema de multicanalización por división en frecuencia emplea la modulación de subportadora de banda lateral superior y modulación de portadora de FM. Hay K señales de entrada independientes, cada una de banda limitada en W_0, y las frecuencias subportadoras son $f_{ck} = (K-1)W_0$, $K = 1, 2, \ldots, K$. La señal de banda base aplicada al modulador de FM es $x_b(t) = \Sigma\, \alpha_K x_K(t)$ donde $x_K(t)$ es la señal subportadora k-ésima, teniendo cada una $x_K{}^2 = 2$, y α_K son constantes. En el receptor se tienen a continuación del discriminador un banco de filtros pasabanda y detectores sincrónicos. Demuéstrese que la S/N a la salida del canal k-ésimo es $(S/N)_K = f_\Delta{}^2\alpha K^2/N_K$ donde $N_K = (3K^2 - 3K + 1)\eta W_0{}^3/3S_R$. (*Sugerencia*: Recuérdese que la S/N en la entrada y en la salida son iguales para detección sincrónica de SSB.)

7.32★ (Secc. 7.6) En el sistema descrito en el problema 7.31 las α_K se han escogido de tal manera que $x_b{}^2 = 1$ y todos los canales tienen el mismo valor de $(S/N)_K$. Obténganse expresiones para α_K y $(S/N)_K$ bajo estas condiciones.

7.33★ (Secc. 7.6) La multicanalización en FM estéreo emplea filtraje de preacentuación/desacentuación que no se muestra en las figuras 5.37 y 5.38. De manera específica, las señales matrizadas $x_1(t) = x_L(t) + x_R(t)$ y $x_2(t) = x_L(t)$ están cada una de ellas preacentuadas antes de ser multicanalizadass en el transmisor; de manera similar, cada una de ellas es desacentuada después de demulticanalizada, pero con anterioridad a la matriz en el receptor. En toda la operación se usan filtros RC con $B_{de} = 2.1$ kHz. Las entradas de señal más ruido a la matriz del receptor son $y_1(t) = f_\Delta x_1(t) + n_1(t)$ y $y_2(t) = f_\Delta x_2(t) + n_2(t)$, y las salidas finales son $y_1(t) \pm y_2(t)$.
(*a*) Demuéstrese que $n_2(t)$ tiene $G_{n_2}(f) = |H_{de}(f)|^2(\eta/S_R)(f^2 + f_{sc}{}^2)\Pi(f/2W)$, donde $f_{sc} = 38$ kHz, suponiendo que el filtro pasabanda y la detección sincrónica son ideales. Después demuéstrese que $n_2{}^2 \gg n_1{}^2$. (*Sugerencia*: Bosquéjese el espectro de ruido en la salida del filtro pasabanda y aplíquese la traslación de frecuencia.)
(*b*) Tomando a $x_L x_R = 0$ y $\overline{x_L{}^2} = \overline{x_R{}^2} \approx ⅓$ —por lo que $\overline{x_b{}^2} \approx 1$ si la amplitud piloto es pequeña— demuéstrese que la $(S/N)_D$ por canal para transmisión en estéreo es alrededor de 20 dB menos que en la transmisión monoaural.

8

MUESTREO Y MODULACIÓN DE PULSOS

Los datos experimentales y las funciones matemáticas con frecuencia se presentan como curvas *continuas*, aunque se puede emplear un número finito de *puntos discretos* para construir la gráfica. Si estos puntos discretos o *muestras*, tienen un espaciamiento suficientemente próximo, se dibuja una curva uniforme uniéndolos, y los valores intermedios se pueden interpolar con un grado razonable de exactitud. Por lo tanto, se puede decir que la presentación continua se describe en forma adecuada por medio de los puntos muestra.

En forma similar, una señal eléctrica que satisface ciertos requisitos, se puede reproducir enteramente con un conjunto apropiado de muestras instantáneas. Si esto se cumple, la teoría del muestreo establecerá las condiciones necesarias y sólo se necesitará transmitir los valores muestra según ocurran en vez de enviar la señal en forma continua. Esto es *modulación de pulsos*.

La clave distintiva entre la modulación de pulsos y la de CW es como sigue: en la modulación de CW, algún parámetro de la onda modulada varía en forma continua con el mensaje; en la modulación de pulsos, algún parámetro de cada pulso se modula por un valor muestra particular del mensaje. Por lo general, los pulsos son muy cortos en comparación con el tiempo entre ellos, por lo que una onda modulada por pulsos está "fuera" la mayor parte del tiempo.

Por esta propiedad, la modulación de pulsos ofrece dos ventajas potenciales sobre la CW: Primera, la potencia transmitida se puede concentrar en ráfagas cortas

en vez de ser enviada en forma continua. Esto le da al ingeniero de sistemas una libertad adicional para la selección del equipo, puesto que ciertos dipositivos, como válvulas de microondas de alta potencia y láseres, son operables sólo en una base de pulsos. Segunda, los intervalos entre pulsos se pueden llenar con valores muestra de otros mensajes, permitiéndose así la transmisión de muchos mensajes en un sistema de comunicación. A tal multicanalización en el dominio del tiempo se le conoce como *multicanalización por división en tiempo* (TDM).

Otra diferencia entre las modulaciones de pulsos y de CW reside en que la onda pulsada puede contener CD y contenido de baja frecuencia apreciables. Una transmisión eficiente incluye, por tanto, una segunda operación; modulación de CW para propiciar una traslación completa de frecuencia. Bajo este punto de vista, la modulación de pulsos es una técnica de *procesamiento del mensaje* más que una modulación en el sentido usual. Efectivamente, el uso más común de la modulación de pulsos es el procesamiento del mensaje empleando la multicanalización por división en tiempo.

Existen dos tipos básicos de modulación de pulsos: la analógica, como la modulación de amplitud de pulsos o la de posición de pulsos, la cual es en muchas formas similar a las modulaciones lineal o exponencial, y la modulación *digital* o de *pulsos codificados*, la cual no tiene equivalencia con la de CW. Ambos tipos se examinan en este capítulo, recalcándose la última en razón a sus propiedades únicas y deseables.

Independientemente del tipo, la operación clave para comunicación pulsada es la extracción de valores muestra de la forma de onda del mensaje. Más aún, el concepto del muestreo juega un papel igualmente importante en la *teoría de la información*, el tema del capítulo 9, y la teoría de los sistemas de control de datos muestreados. Por ello, se empieza este capítulo con un estudio del muestreo, tanto en forma teórica, como práctica.

8.1 TEORÍA Y PRÁCTICA DEL MUESTREO

Una aproximación simple pero bastante informativa a la teoría del muestreo, se consigue por medio de la operación de conmutación de la figura 8.1*a*. El conmutador se

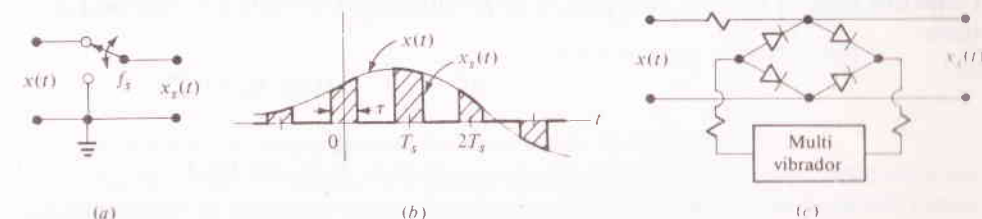



FIGURA 8.1
Un muestreador de conmutación. (*a*) Diagrama; (*b*) formas de onda; (*c*) circuito electrónico.

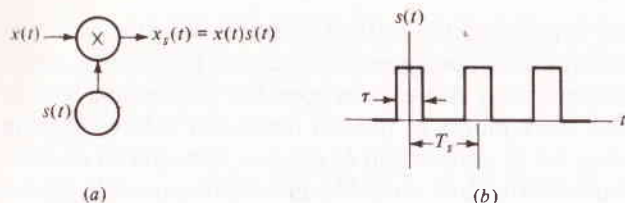



FIGURA 8.2
El muestreo interpretado como multiplicación. (*a*) Diagrama; (*b*) función de conmutación $s(t)$.

desplaza en forma periódica entre los dos contactos a una velocidad de $f_s = 1/T_s$, permaneciendo en el contacto de la señal de entrada durante τ segundos y en el contacto conectado a tierra durante lo que resta de cada periodo. La salida $x_s(t)$ está formada entonces de segmentos cortos de la entrada $x(t)$, como se muestra en la figura 8.1*b*. La figura 8.1*c* es una versión electrónica de la figura 8.1*a*; el voltaje de salida es igual al voltaje de entrada excepto cuando el multivibrador polariza en sentido directo a los diodos y con ello se lleva a la salida a cero. Esta operación, a la cual se le designa algunas veces como *de un solo terminal o troceado unipolar*, no es un muestreo instantáneo en el estricto sentido. No obstante, a $x_s(t)$ se le designará como la onda muestreada y a f_s como la frecuencia de muestreo.

Surge ahora la pregunta: ¿Son suficientes los segmentos muestreados para describir a la señal de entrada original?, y de ser así, ¿cómo se puede recuperar a $x(t)$ de $x_s(t)$? La respuesta a estas preguntas cae en el dominio de la frecuencia, o sea, en el espectro de la onda muestreada.

Como un primer paso en la búsqueda del espectro, se introduce una *función de conmutación* $s(t)$ tal que

$$x_s(t) = x(t)s(t) \tag{1}$$

Así, la operación de muestreo viene a ser una multiplicación por $s(t)$, como se indica en forma esquemática en la figura 8.2*a*, donde $s(t)$ no es otra cosa que el tren periódico de pulsos de la figura 8.2*b*. Puesto que $s(t)$ es periódico, se le puede escribir como una serie de Fourier. Empleando los resultados del ejemplo 2.2, sección 2.2, se tiene

$$\begin{aligned} s(t) &= \sum_{n=-\infty}^{\infty} f_s \tau \operatorname{senc} n f_s \tau \, e^{j2\pi n f_s t} \\ &= c_0 + \sum_{n=1}^{\infty} 2c_n \cos n\omega_s t \end{aligned} \tag{2}$$

donde

$$c_n = f_s \tau \operatorname{senc} n f_s \tau \qquad \omega_s = 2\pi f_s$$

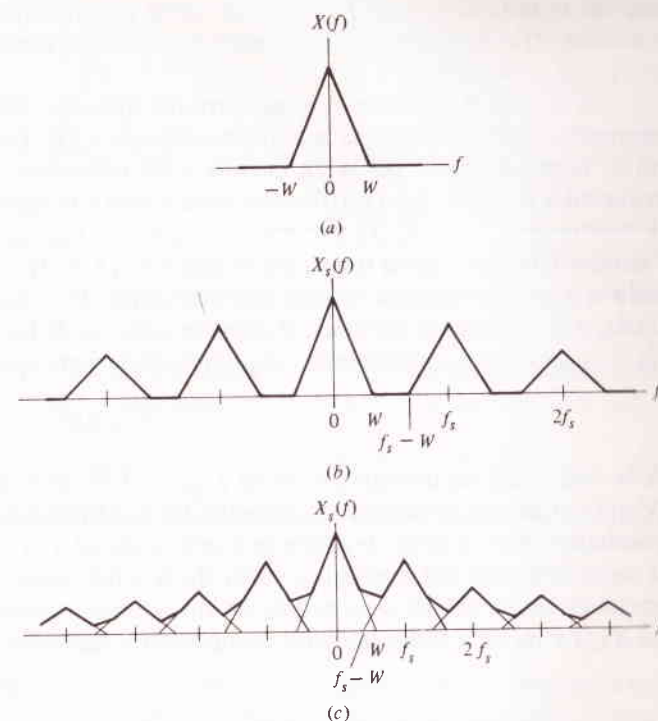



FIGURA 8.3
Espectros para muestreo de conmutación. (*a*) Mensaje; (*b*) mensaje muestreado, $f_s > 2W$; (*c*) mensaje muestreado, $f_s < 2W$.

Combinando la ecuación (2) con la (1) se llega al desarrollo término a término

$$x_s(t) = c_0 x(t) + 2c_1 x(t) \cos \omega_s t + 2c_2 x(t) \cos 2\omega_c t + \cdots \tag{3}$$

Así, si el espectro de entrada es $X(f) = \mathscr{F}[x(t)]$, el espectro de salida es

$$\begin{aligned} X_s(f) = c_0 X(f) &+ c_1[X(f-f_s) + X(f+f_s)] \\ &+ c_2[X(f-2f_s) + X(f+2f_s)] \\ &+ \cdots \end{aligned} \tag{4}$$

lo cual procede en forma directa del teorema de la modulación.

En tanto que la ecuación (4) se presenta bastante confusa, el espectro de la onda muestreada se bosqueja con facilidad si se supone que la señal es de *banda limitada*, digamos en W. La figura 8.3 muestra una $X(f)$ conveniente y la correspondiente $X_s(f)$ para dos casos, $f_s > 2W$ y $f_s < 2W$. Examinando esta figura se revela algo muy sorprendente: la operación de muestreo ha dejado al espectro del mensaje *intacto*, repitiéndolo simplemente en forma periódica en el dominio de la frecuencia

con un espaciamiento de f_s. También debe notarse que el primer término de la ecuación (4) es precisamente el espectro del mensaje, atenuado por el *ciclo de trabajo* $c_0 = f_s\tau = \tau/T_s$.

Si el muestreo conserva al espectro del mensaje, debe ser posible recuperar o *reconstruir* a $x(t)$ a partir de la onda muestreada $x_s(t)$. La técnica de reconstrucción no es completamente obvia en cuanto a las relaciones, del dominio del tiempo, ecuaciones (1) y (3), pero refiriéndonos de nuevo a la figura 8.3, se ve que a $X(f)$ se le puede separar de $X_s(f)$ por medio de *filtraje pasabajas*, siempre y cuando las "bandas laterales" no se traslapen. Y si se filtra a $X(f)$ sola de $X_s(f)$, se ha recuperado a $x(t)$. De manera obvia, son necesarias dos condiciones para prevenir el traslape de las bandas laterales: el mensaje debe ser de banda limitada, y la frecuencia de muestreo lo suficientemente grande como para que $f_s - W \geq W$, esto es,

$$f_s \geq 2W \qquad \text{o} \qquad T_s \leq \frac{1}{2W} \tag{5}$$

A la frecuencia de muestreo mínima $f_{s_{\min}} = 2W$ se le conoce como *velocidad de Nyquist*. Cuando se satisface la ecuación (5) y se filtra a $x_s(t)$ por medio de un filtro pasabajas ideal, la señal de salida será proporcional a $x(t)$; por lo que se ha logrado la reconstrucción del mensaje a partir de la señal muestreada. El valor exacto del ancho de banda del filtro no es dato de importancia mientras el filtro permita el paso de $X(f)$ y rechace todas las otras componentes, es decir,

$$W \leq B \leq f_s - W \tag{6}$$

como se desprende de la figura 8.3*b*.

Este análisis ha mostrado que si una señal de banda limitada se muestrea a una frecuencia mayor que la velocidad de Nyquist, se le puede *reconstruir en forma completa* a partir de la onda muestreada. La reconstrucción se concluye empleando filtraje de pasabajas. Estas conclusiones pueden resultar difíciles de creer por parte del lector a la primera exposición; pues ciertamente ponen a prueba la fe que se tenga al análisis espectral. No obstante, son bastante correctas.

Por último, se debe puntualizar que estos resultados son independientes de la duración del pulso muestra, salvo cuando aparece en el ciclo de trabajo. Si se hace a τ muy pequeña, $x_s(t)$ se aproxima a una cadena de *puntos muestra instantáneos*, lo cual corresponde al *muestreo ideal*. Y mientras el análisis anterior sea representativo del muestreo práctico y sea suficiente para muchos propósitos, la atención deberá enfocarse hacia el muestreo ideal, el estudio del cual permite llegar a un mejor conocimiento del cómo y porqué del muestreo y la reconstrucción; además de la prueba del teorema del muestreo uniforme.

EJERCICIO 8.1 Considérese el caso de un troceador imperfecto donde los pulsos que conforman a $s(t)$ no son rectangulares, por lo que

$$s(t) = \sum_{k=-\infty}^{\infty} p(t - kT_s) \tag{7a}$$

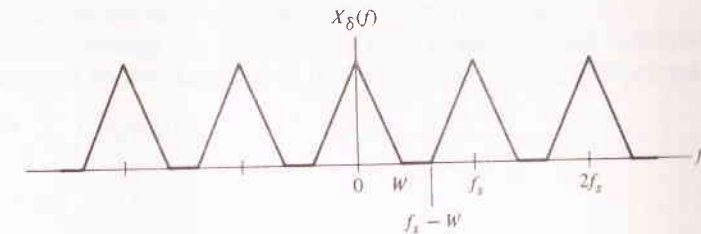



FIGURA 8.4
Espectro de un mensaje muestreado en forma ideal.

donde la forma de pulso $p(t)$ es igual a cero para $|t| > T_s/2$ pero que de otra manera es arbitraria. Haciendo $P(f) = \mathscr{F}[p(t)]$ y $c_n = P(nf_s)/T_s$, demuéstrese que

$$x_s(t) = c_0\, x(t) + \sum_{n=1}^{\infty} |2c_n|\, x(t) \cos(n\omega_s t + \arg[c_n]) \tag{7b}$$

y en consecuencia las conclusiones anteriores relativas a la recuperación de $x(t)$ serán siempre válidas.

Muestreo y reconstrucción ideales

Por definición, el muestreo ideal es el *instantáneo*. El dispositivo de conmutación de la figura 8.1*a* produce valores instantáneos sólo si $\tau \to 0$, y así lo hace $x_s(t)$ —véase la ecuación (3)—. De manera conceptual, esta dificultad se supera multiplicando a $x_s(t)$ por $1/\tau$ de manera que, conforme $\tau \to 0$ y $1/\tau \to \infty$, la onda muestreada resulta ser un tren de *impulsos* cuyas *áreas* son iguales a los valores muestra instantáneos de la señal de entrada. De manera formal se define

$$s_\delta(t) \triangleq \sum_{k=-\infty}^{\infty} \delta(t - kT_s) \tag{8}$$

expresión que se conoce como la función de muestreo ideal. Entonces, en la notación del ejercicio 8.1, $p(t) = \delta(t)$, $P(f) = \mathscr{F}[\delta(t)] = 1$, y $c_n = 1/T_s = f_s$ para toda n. Por lo que, de la ecuación (7*b*),

$$x_\delta(t) \triangleq x(t)s_\delta(t) \tag{9a}$$

$$= f_s\left[x(t) + \sum_{n=1}^{\infty} 2x(t)\cos n\omega_s t\right] \tag{9b}$$

y

$$X_\delta(f) = \mathscr{F}[x_\delta(t)]$$

$$= f_s \sum_{n=-\infty}^{\infty} X(f - nf_s) \tag{10}$$

lo cual se ilustra en la figura 8.4 como el espectro del mensaje de la figura 8.3*a*, tomando $f_s > 2W$. Nótese que $X_\delta(f)$ es *periódica en frecuencia* con periodo f_s, una observación decisiva en el estudio de los sistemas de datos muestreados.

A $x_\delta(t)$ se le denomina como la *onda muestreada ideal*; consiste en un tren de impulsos con espaciamiento T_s cuyas áreas son iguales a los valores muestra instantáneos $x(kT_s)$. Esto se expresa con la combinación de las ecuaciones (8) y (9*a*), así:

$$x_\delta(t) = x(t) \sum_{k=-\infty}^{\infty} \delta(t - kT_s) \qquad (11a)$$

$$= \sum_{k=-\infty}^{\infty} x(kT_s)\, \delta(t - kT_s) \qquad (11b)$$

puesto que $x(t)\delta(t - t_d) = x(t_d)$.

Entre paréntesis, se puede desarrollar también una expresión para $S_\delta(f) = \mathscr{F}[S_\delta(t)]$ como sigue. De la ecuación (9*a*) y del teorema de la convolución, $X_\delta(f) = [X(f)] * [S_\delta(f)]$ mientras que la ecuación (10) es equivalente a

$$X_\delta(f) = [X(f)] * \left[\sum_{n=-\infty}^{\infty} f_s\, \delta(f - nf_s)\right]$$

Por lo cual, se concluye que

$$S_\delta(f) = f_s \sum_{n=-\infty}^{\infty} \delta(f - nf_s) \qquad (12)$$

así, el espectro de una cadena periódica de impulsos de peso unitario en el dominio del tiempo, es periódica de impulsos en el dominio de la frecuencia con espaciamiento $f_s = 1/T_s$; en ambos dominios se tiene una función que luce como una cerca de puntas.

Volviendo al tema principal y a la figura 8.4, se ve de manera inmediata que si se recurre a las mismas condiciones de antes —$x(t)$ de banda limitada en W y $f_s \geq 2W$— entonces con un filtro de ancho de banda adecuado se reconstruirá a $x(t)$ a partir de la onda muestreada ideal. De manera específica para un filtro pasabajas ideal de ganancia de voltaje K, con retardo en el tiempo t_d, y ancho de banda B, la función de transferencia es

$$H(f) = K\Pi\left(\frac{f}{2B}\right)e^{-j\omega t_d}$$

por lo que filtrando a $x_s(t)$ se produce el espectro de salida

$$Y(f) = H(f)X_\delta(f) = Kf_s X(f)e^{-j\omega t_d}$$

suponiendo que B satisface la ecuación (6). La función del tiempo de salida es entonces

$$y(t) = \mathscr{F}^{-1}[Y(f)] = Kf_s x(t - t_d) \qquad (13)$$

la cual es la señal original amplificada en Kf_s, retardada en t_d.

Se puede adquirir mayor confianza en el proceso de muestreo examinando la reconstrucción en el dominio del tiempo. La respuesta de impulso del filtro anterior es

$$h(t) = 2BK \operatorname{senc} 2B(t - t_d)$$

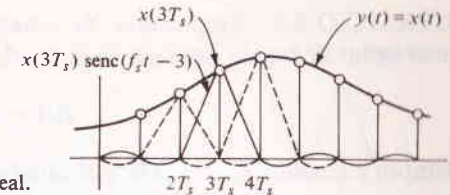



FIGURA 8.5
Reconstrucción ideal.

Y puesto que la entrada $x_\delta(t)$ es un tren de impulsos ponderados, la salida es un tren de *respuestas de impulso* ponderados, a saber,

$$y(t) = [h(t)] * [x_\delta(t)] = \sum_k x(kT_s)h(t - kT_s)$$

$$= 2BK \sum_{k=-\infty}^{\infty} x(kT_s) \operatorname{senc} 2B(t - t_d - kT_s) \qquad (14)$$

Ahora supóngase para efectos de sencillez que $B = f_s/2$, $K = 1/f_s$ y $t_d = 0$, por lo que

$$y(t) = \sum_k x(kT_s) \operatorname{senc}(f_s t - k)$$

Se puede desarrollar entonces el proceso de reconstrucción en forma gráfica, como se muestra en la figura 8.5. Es claro que los valores correctos se reconstruyen en los instantes de muestreo $t = kT_s$, porque todas las funciones senc son cero a estos tiempos salvo una, que es en la que se encuentra $x(kT_s)$. Entre los instantes de muestreo se *interpola* a $x(t)$ sumando los precursores y los postcursores de *todas* las funciones senc. Por esta razón, al filtro pasabajas a menudo se le designa *filtro de interpolación*, y a su respuesta de impulso *función de interpolación*.

El teorema del muestreo uniforme

Los resultados anteriores se resumen bien estableciendo el importante teorema del muestreo uniforme (periódico). Mientras existan muchas variantes de este teorema,† su forma siguiente es la más adecuada para nuestros propósitos.

> Si una señal no contiene componentes de frecuencia para $|f| \geq W$, se describe en forma completa por medio de valores muestra instantáneos espaciados de manera uniforme en el tiempo con periodo $T_s \leq 1/2W$. Si una señal ha sido muestreada a la velocidad de Nyquist o a una mayor ($f_s \geq 2W$) y a los valores muestra se les representa como impulsos ponderados, la señal puede ser reconstruida en forma exacta a partir de sus muestras por medio de un filtro pasabajas ideal de ancho de banda B, donde $W \leq B \leq f_s - W$.

† Black (1953, Cap. 4) lista varios teoremas que se refieren al muestreo periódico.

EJERCICIO 8.2 Empleando las ecuaciones (13) y (14), demuéstrese que a cualquier señal de banda limitada se le puede desarrollar en una serie de funciones senc

$$x(t) = 2BT_s \sum_{k=-\infty}^{\infty} x(kT_s) \operatorname{senc} 2B(t - kT_s) \tag{15}$$

siempre y cuando $T_s \leq 1/2W$ y B satisfaga la ecuación (6). Por lo que, en la misma forma que una señal periódica se describe en forma completa por medio de sus coeficientes de serie de Fourier, a una señal de banda limitada se le describe en forma completa por medio de sus valores muestra instantáneos *sea que a la señal se le muestree en realidad o no.*

Muestreo práctico

Habiendo establecido y probado el teorema para el muestreo ideal, se debe examinar ahora el muestreo para la forma de onda eléctrica como ocurre en la práctica. El muestreo práctico difiere del ideal en tres aspectos obvios:

1 La onda muestreada está formada de pulsos que tienen amplitud y duración finitas, al contrario de los impulsos.
2 Los filtros de reconstrucción prácticos no son filtros ideales.
3 Los mensajes sujetos a muestreo son *limitados en tiempo* y por ello no pueden ser de banda limitada.

Se tratarán estas diferencias una a una y se demostrará que sólo la última resulta problemática.

Considerando los efectos de la forma de los pulsos, la investigación del troceador unipolar y del ejercicio 8.1 ha demostrado que en forma virtual cualquier forma de pulso $p(t)$ funcionará cuando la operación de muestreo se modele en forma matemática como multiplicación, $x_s(t) = x(t)S(t)$. Algunos otros dispositivos de muestreo, en forma especial la variedad de muestreo y retención, generan una onda muestreada de la forma

$$x_s(t) = \sum_k x(kT_s)p(t - kT_s)$$

$$= [p(t)] * \left[\sum_k x(kT_s)\delta(t - kT_s)\right] \tag{16a}$$

de manera que

$$X_s(f) = P(f)\left[f_s \sum_n X(f - nf_s)\right] = P(f)X_\delta(f) \tag{16b}$$

donde $P(f) = \mathscr{F}[p(t)]$ y $X_\delta(f)$ es el espectro de la onda muestreada ideal.

Si interpretamos la ecuación (16*b*), se puede pensar que $P(f)$ es un filtro operando sobre $X_\delta(f)$ y atenuando a todas las componentes de frecuencia superiores a $1/\tau$, siendo τ la duración nominal de $p(t)$. Por lo general, la señal reconstruida estará distorsionada, dado que $P(f)$ atenúa la porción superior del espectro del mensaje. A esta pérdida de los componentes de alta frecuencia del mensaje se le designa algu-

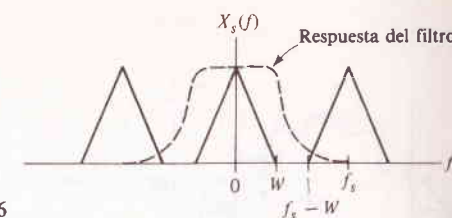



FIGURA 8.6

nas veces como *efecto de apertura* —la apertura del muestreo (duración de pulso) es muy grande— y se puede corregir por medio de un *filtro de igualación* que tenga $H_{eq}(f) = 1/P(f)$. Sin embargo, por lo general $1/\tau \gg W$, por lo que $P(f)$ es esencialmente constante sobre la banda del mensaje y la igualación puede no requerirse. Se puede así decir que los efectos de forma del pulso son inconsecuentes de manera relativa, y el teorema del muestreo resulta válido para ondas muestreadas no impulsivas.

El efecto de los filtros de reconstrucción no ideales se trata también con facilidad en el dominio de la frecuencia. Considérese, por ejemplo, una respuesta típica de filtro superpuesta a un espectro de onda muestreada, como se ilustra en la figura 8.6. Si el filtro es plano en forma razonable sobre la banda del mensaje, su salida consistirá de $x(t)$ más componentes espurias de frecuencia en $|f| > f_s - W$, lo cual está fuera de la banda del mensaje. Nótese que estas componentes están atenuadas en forma considerable en comparación con $x(t)$. En sistemas de audio, tales componentes se escucharían como mutilación de alta frecuencia o silbidos. Sin embargo, su intensidad es proporcional al mensaje, desapareciendo cuando $x(t) = 0$, y el mensaje tiende a enmascararles, haciéndolos más fáciles de tolerar.

Obviamente, el mejor procedimiento para reducir las frecuencias espurias es el diseño de un buen filtro. En forma alterna, para una respuesta de filtro dada, éstas se pueden suprimir mejor incrementando la frecuencia de muestreo, la cual crece en $f_s - W$ y producen *bandas de seguridad* en el espectro de la onda muestreada. Hay en esto una semejanza interesante con la multicanalización por división en frecuencia: en ésta, las bandas de seguridad se emplean para permitir la separación del mensaje por medio de filtros pasabanda prácticos con un mínimo de *diafonía*; en los sistemas de muestreo, las bandas de seguridad se emplean para permitir la reconstrucción del mensaje por medio de filtros pasabajas prácticos con un mínimo de mutilación de alta frecuencia.

Ejemplo 8.1 Reconstrucción mediante el uso de una retención de orden cero.

Cuando la señal muestreada está formada de pulsos rectangulares planos, se consigue una aproximación a una escalera simplemente extendiendo los pulsos en $x(t)$ como se muestra en la figura 8.7*a*. En forma analítica, haciendo $h(t) = \Pi[(t - \tfrac{1}{2}T_s)/T_s]$, la salida es

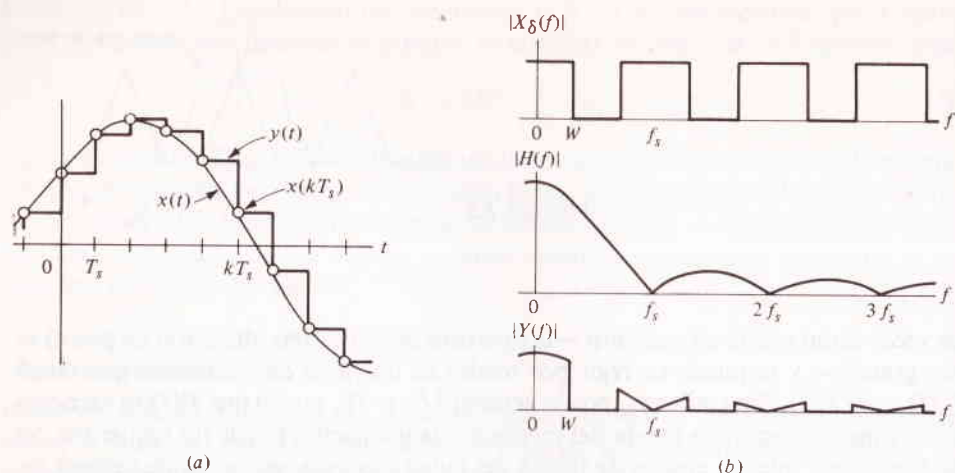



FIGURA 8.7
Reconstrucción empleando una retención de orden cero. (*a*) Forma de onda de escalera; (*b*) espectros.

$$y(t) = \sum_k x(kT_s)h(t - kT_s) = [h(t)] * [x_\delta(t)]$$

y

$$Y(f) = H(f)X_\delta(f) \qquad H(f) = \mathscr{F}[h(t)]$$

expresión similar a las ecuaciones (16*a*) y (16*b*). Pero $h(t)$ tiene la misma forma que la respuesta de impulso de la *retención de orden cero*† descrita (Ej. 2.13, Sec. 2.5). Por tanto, $|H(f)| = T_s|\operatorname{senc} fT_s|$ y $|y(f)|$ es como se bosqueja en la figura 8.7*b*, considerando a $X(f)$ constante sobre $|f| \leq W$.

Es obvio que a este método de reconstrucción le afecta tanto el efecto de apertura como el "ruido" fuera de banda, los cuales se pueden disminuir por medio de filtraje adicional. Si, no obstante, $f_s \gg 2W$, ambos efectos son pequeños y $y(t)$ es una buena aproximación a $x(t)$. La condición $f_s \gg 2W$ no es muy práctica para transmisión de señal, pero lo es para los sistemas de control de datos muestreados y simulación digital donde se emplea comúnmente la reconstrucción de escalera.

Señales de banda no limitada. Interferencia de colas espectrales

Hasta ahora hemos tomado en cuenta el hecho de que las señales reales no son de banda estrictamente limitada. Pero un espectro de mensaje como el de la figura 8.8*a* se considera que, en forma virtual, es de banda limitada si el contenido de frecuencia

† El nombre proviene de la propiedad de que una retención de orden *m*-ésimo puede reconstruir de manera perfecta señales de la forma $x(t) = a_0 + a_1t + \ldots + a_mt^m$.

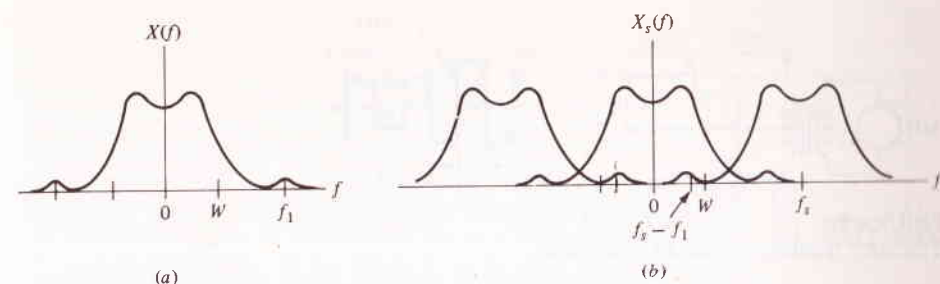



FIGURA 8.8
Efecto de interferencia de colas espectrales: (*a*) Espectro de mensaje de banda no limitada; (*b*) espectro de mensaje muestreado con traslapes.

arriba de W es pequeño y se presume que no es importante en lo que a conducción de información se refiere. Cuando se muestrea tal mensaje, habrá un traslape inevitable de componentes espectrales, figura 8.8*b*. En la reconstrucción, las frecuencias originalmente fuera de la banda nominal del mensaje aparecerán a la salida del filtro en la forma de frecuencias mucho *menores*. Así, por ejemplo, $f_1 > W$ resulta ser $f_s - f_1 < W$, como se indica en la figura.

Este fenómeno de traslación de frecuencia hacia valores menores ocurre siempre que se *submuestrea* una componente de frecuencia, o sea que, $f_s < 2f$, y se le da el nombre descriptivo de *interferencia de colas espectrales (aliasing)*. El efecto de esta interferencia es más serio que el de las frecuencias espurias que pasan por filtros de reconstrucción no ideales, ya que las últimas caen *fuera* de la banda del mensaje, mientras que las componentes interferentes pueden caer *dentro* de ésta. La interferencia de colas espectrales se remedia filtrando el mensaje tanto como sea posible *antes* del muestreo y, si es necesario, muestreando a una velocidad mucho mayor que la nominal de Nyquist.

Como ilustración, el espectro de voz promedio (Fig. 5.6*b*) se extiende más allá de los 10 kHz, aunque mucha de la energía se concentra en el intervalo de 100 a 600 Hz y un ancho de banda de 3 kHz es suficiente para inteligibilidad. Cuando una onda de voz se procesa primero por medio de un filtro pasabajas de 3 kHz y después se muestrea a $f_s = 8$ kHz, lo normal para sistemas telefónicos, las componentes interferentes de colas espectrales están en forma típica 30 dB debajo de la señal deseada y en forma virtual no se perciben.

EJERCICIO 8.3 Demuéstrese el efecto de interferencia de colas espectrales haciendo un bosquejo cuidadoso de $\cos 2\pi 10t$ y $\cos 2\pi 70t$ para $0 \leq t \leq 1/10$. Pónganse ambos bosquejos sobre el mismo conjunto de ejes y encuéntrense los valores muestra en $t = 0, 1/80, 2/80, \ldots, 8/80$, lo cual corresponde a $f_s = 80$. También debe convencerse el lector de que no se puede interpolar ninguna otra forma de onda de banda limitada en $10 < W < 40$ a partir de los valores muestra de $\cos 2\pi 10t$

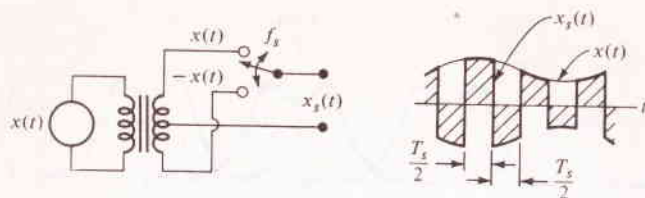



FIGURA 8.9
Troceador bipolar y formas de onda típicas.

El teorema del muestreo reenunciado

Para resumir la investigación sobre el muestreo *práctico* conforme se diferenció del muestreo *ideal*, se puede reenunciar el teorema del muestreo de la siguiente manera:

> Si una señal se ha filtrado a pasabajas dando lugar a un contenido espectral despreciable para $|f| > W$, ella queda descrita en una forma adecuada para muchos propósitos por medio de valores muestra, ya sean instantáneos o de duración finita, espaciados de manera uniforme en el tiempo con periodo $T_s \leq 1/2W$. Si una señal ha sido muestreada a la velocidad nominal de Nyquist o mayor ($f_s \geq 2W$) y las muestras se representan por pulsos periódicos cuyas amplitudes son proporcionales a los valores muestra, la señal se puede reconstruir en forma aproximada a partir de sus muestras por medio de filtraje de pasabajas.

De acuerdo en que la reconstrucción *perfecta*, como la transmisión sin distorsión, es algo imposible; pero con una planeación cuidadosa, un buen diseño de filtros y una frecuencia de muestreo adecuada, la onda recuperada se puede hacer tan parecida a la señal original como se desee para todos los casos, excepto para los muy extremos.

Modulación de troceador

Antes de aplicar la teoría del muestreo a los sistemas de modulación de pulsos, se debe puntualizar que los muestreadores prácticos se pueden emplear como *moduladores balanceados* para los sistemas de modulación lineal. En particular, regresando al espectro de salida del troceador unipolar (Fig. 8.3*b*), se ve que cada armónica de la frecuencia de muestreo está *modulada a doble banda lateral* por la señal de entrada $x(t)$. El filtraje pasabanda de la onda muestreada producirá así modulación de doble banda lateral, una técnica conocida como *modulación de troceador*. La modulación no se afecta en forma relativa por una conmutación imperfecta, como se demostró al estudiar los efectos de la forma de pulso.

Para los propósitos de la modulación de portadora suprimida, los métodos de conmutación o troceado son a menudo superiores al del modulador balanceado convencional, el cual requiere de más componentes y de elementos no lineales acoplados en forma cuidadosa. Sin embargo, el troceador unipolar no es un modulador tan eficiente como el troceador *bipolar* de la figura 8.9. Este dispositivo produce modulación de doble banda lateral sólo a las armónicas pares de f_s, y su salida no contiene componentes de CD. El análisis procede del realizado para el troceador unipolar, pero con una función de conmutación de onda cuadrada.

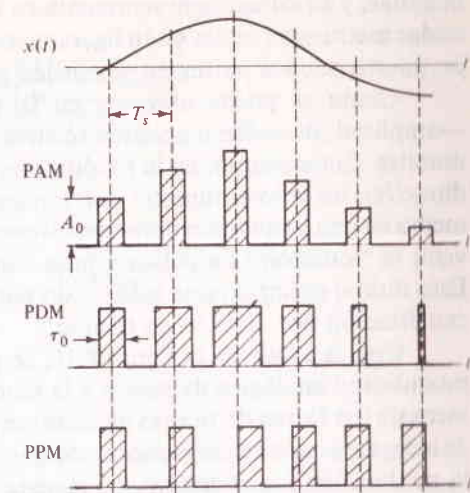



FIGURA 8.10
Tipos de modulación analógica de pulsos.

Aparte de la modulación para transmisión de señales, los troceadores bipolares encuentran aplicación en sistemas de control, donde se emplea una onda con portadora suprimida para impulsar servomotores bifásicos, y para la amplificación de señales que varían con lentitud (amplificadores de CD). De manera adicional, el problema 8.12 indica cómo se puede modificar un troceador bipolar para lograr la multicanalización de banda base en el caso de los transmisores de FM estereofónicos.

8.2 MODULACIÓN ANALÓGICA DE PULSOS: MODULACIÓN EN AMPLITUD DE PULSOS (PAM), MODULACIÓN EN DURACIÓN DE PULSOS (PDM), Y MODULACIÓN EN POSICIÓN DE PULSOS (PPM)

Si un mensaje se describe en forma adecuada por sus valores muestra, se le puede transmitir por medio de la modulación analógica de pulsos, en donde los valores muestra modulan en forma directa a un tren periódico de pulsos con un pulso de cada muestra. Existen numerosas variedades de modulación analógica de pulsos, y la terminología no se ha estandarizado. Sin embargo, los tres tipos que se analizarán aquí se designan por lo general como *modulación en amplitud de pulsos* (PAM), *modulación en duración de pulsos* (PDM) y *modulación en posición de pulsos* (PPM). Los dos últimos también se agrupan juntos bajo el encabezado general de *modulación de pulsos en tiempo*.

La figura 8.10 muestra un mensaje típico y las correspondientes formas de onda de los pulsos modulados. Para claridad, los pulsos se muestran en forma rec

tangular, y su duración se representa en forma tosca por lo exagerada. Más aún, las ondas moduladas reales están ligeramente retardadas en comparación con el mensaje, puesto que los pulsos no se pueden generar antes de los instantes de muestreo.

Como se puede observar en la figura, el parámetro de pulso modulado —amplitud, duración o posición relativa— varía en proporción directa a los valores muestra. Sin embargo, en la modulación de pulsos en amplitud y en la de pulsos en duración, los valores muestra que son iguales a cero se representan casi siempre por medio de una amplitud o duración diferentes de cero. Se sigue esta práctica para prevenir la "omisión" de pulsos y para conservar una velocidad de pulsos constante. Esto último es importante sobre todo para propósitos de sincronización en la multicanalización por división en tiempo.

Con la ayuda de la figura 8.10, se pueden obtener ciertas semejanzas entre la modulación analógica de pulsos y la modulación de CW. Así, a la información del mensaje (en forma de valores muestra) se le conduce en la *amplitud* de una onda modulada en amplitud de pulsos, por lo que este tipo de modulación es análogo a la modulación *lineal* de CW, de manera particular a la AM. En la modulación en duración de pulsos y en posición de pulsos, a la información se le conduce en un parámetro *tiempo*, o sea, la ubicación de los bordes del pulso. Recordando que la frecuencia y la fase instantáneas son también parámetros de tiempo, se puede decir que la modulación en tiempo de pulsos es análoga a la modulación *exponencial* de CW.

Sistemas de modulación de pulsos

Antes de proceder a un análisis más detallado, se pueden establecer tres observaciones generales acerca de la modulación analógica de pulsos.

1 Las ondas moduladas por pulsos tienen CD y un contenido de baja frecuencia apreciables, en especial cerca de las primeras armónicas de $f_s = 1/T_s$. Por lo que la transmisión directa puede ser difícil, si no que imposible.

2 Cuando se transmiten las ondas moduladas por pulsos, se debe tener cuidado en prevenir el traslape de los pulsos, ya que éste destruiría la modulación. Dados los requisitos de resolución de pulsos de la sección 4.4, se necesita un ancho de banda de banda base no menor a $1/2\tau$, siendo τ la duración nominal del pulso.

3 Las ondas moduladas por pulsos se pueden demodular mediante la *reconstrucción*. De manera conceptual, los valores muestra se extraen de la onda modulada, convertidos en impulsos ponderados, y filtrados a pasabajas.

Con respecto al primer punto, la transmisión de pulsos a corta distancia puede ser factible sobre circuitos alámbricos o por cable coaxial, pero una transmisión eficiente requiere una traslación de frecuencia adicional. En consecuencia, muchos sistemas de pulsos tienen un paso de modulación de portadora, en el cual los pulsos se convierten a pulsos de RF. En forma analítica, si $x_p(t)$ es la onda modulada por pulsos, la señal transmitida efectiva es $x_c(t) = x_p(t)A_c \cos\omega_c t$, donde $f_c \gg f_s$. Aunque la modulación de portadora se clasifica en forma apropiada como modulación de

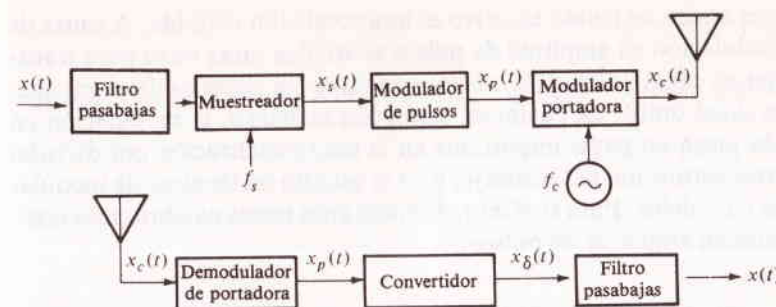



FIGURA 8.11
Sistema de comunicación de modulación de pulsos.

doble banda lateral, se puede emplear la detección de envolvente en el receptor si $x_p(t) \geq 0$ y si no ocurren inversiones de fase en la portadora.

En la figura 8.11 se muestra el diagrama de un sistema completo de transmisión de pulsos. El transmisor está constituido de un filtro pasabajas, un muestreador, un modulador de pulsos y un modulador de portadora. El receptor incluye la demodulación de la portadora seguida de un conversor, el cual cambia la onda modulada por pulsos en un tren de impulsos ponderados $x_\delta(t)$ o en el equivalente realizable. La recuperación del mensaje se efectúa después por medio de filtraje de pasabajas.

En cuanto al ancho de banda de transmisión, las ventajas prácticas de la modulación analógica de pulsos dependen de la duración de éstos, siendo pequeña en comparación con el tiempo entre ellos, o sea

$$\tau \ll T_s \leq \frac{1}{2W} \tag{1}$$

por lo que el ancho de banda de transmisión en banda base es

$$B_{T_{BB}} \geq \frac{1}{2\tau} \gg W \tag{2}$$

y la modulación de portadora duplica esto, dando

$$B_{T_{RF}} = 2B_{T_{BB}} \gg 2W \tag{3}$$

Pero, con o sin modulación de portadora, el ancho de banda que se requiere es grande en comparación con el ancho de banda del mensaje W.

Por analogía con la modulación exponencial, es de sospecharse que del gran ancho de banda de la modulación analógica de pulsos resulta una *reducción de ruido de banda ancha*. Y la modulación *de pulsos en tiempo* en realidad tiene esta útil propiedad. Además, el ancho de banda se determina esencialmente por la magnitud de la reducción de ruido deseada más que por el ancho de banda del mensaje en sí.

Por otra parte, como se podría inferir de su similitud con la AM, la modulación en amplitud de pulsos no es mejor que la transmisión de banda base en cuanto a

ruido se refiere, y el ancho de banda excesivo es una condición definida. A causa de este defecto, la modulación en amplitud de pulsos se emplea raras veces para transmisión de mensajes en canal único. (En efecto, son raros los sistemas de comunicación de pulsos en canal único, de cualquier tipo.) Sin embargo, la modulación en amplitud de pulsos juega un papel importante en la multicanalización por división en tiempo, en la reconstrucción del mensaje, y en el estudio de técnicas de modulación de pulsos más complejos. Para el efecto, diremos unas pocas palabras más acerca de la modulación en amplitud de pulsos.

Modulación en amplitud de pulsos (PAM)

La forma de onda usual de la modulación en amplitud de pulsos está compuesta de pulsos unipolares no rectangulares cuyas amplitudes pico son proporcionales a los valores muestra instantáneos del mensaje. Concretamente,

$$x_p(t) = \sum_k A_0[1 + mx(kT_s)]p(t - kT_s) \tag{4}$$

donde A_0 es la amplitud del pulso no modulada, $p(t)$ es la forma del pulso y m es el índice de modulación comparable en forma directa con el índice de modulación de AM. La condición $[1 + mx(kT_s)] > 0$ se impone en forma normal para conservar la polaridad única y para prevenir la omisión de pulsos, como se mencionó antes. Por lo tanto, recurriendo a la convención de normalización $|x(t)| \leq 1$, el índice de modulación está limitado por $m < 1$.

Puesto que la ecuación (4) tiene la misma forma que la ecuación (16), (Sec. 8.1), con $x(kT_s)$ sustituida por $A_0 + A_0 mx(kT_s)$, se desprende que el espectro lo da la ecuación (16*b*) más impulsos en CD, f_s, $2f_s$, etc. Por lo que un demodulador para la modulación en amplitud de pulsos consiste en forma simple de igualación y filtraje pasabajas con remoción del término de CD, el cual también remueve a $\bar{x}$ si el mensaje tiene una componente de CD.

Para calcular la potencia transmitida promedio $S_T = \langle x_p^2(t)\rangle$, se sigue la aproximación empleada para señales digitales de la sección 4.5. Tomando a $p(t)$ como un pulso rectangular de duración τ, para sencillez, la energía en el pulso k-ésimo es $A_0^2[1 + mx(kT_s)]^2\tau$ por lo que

$$S_T = \frac{1}{T_s} A_0^2 \langle [1 + mx(kT_s)]^2 \rangle \tau$$

$$= \frac{A_0^2 \tau}{T_s}(1 + m^2\overline{x^2}) \tag{5}$$

suponiendo que $\bar{x} = 0$.

Modulación en duración de pulsos (PDM)

En la modulación en duración de pulsos, la duración del pulso k-ésimo conduce al valor muestra. En forma matemática,

$$\tau(k) = \tau_0[1 + mx(kT_s)] \tag{6a}$$

donde τ_0 es la duración no modulada y m juega el mismo papel que antes. Así

$$x_p(t) = \sum_k A_0 p\left[\frac{t - kT_s}{\tau(k)/\tau_0}\right] \tag{6b}$$

y

$$S_T = \frac{A_0^2}{T_s}\langle \tau(k) \rangle = \frac{A_0^2 \tau_0}{T_s} \tag{7}$$

de nuevo tomando a $\bar{x} = 0$ y siendo $p(t)$ rectangular.

La ecuación (6) representa una onda de modulación en duración de pulsos en la cual están modulados ambos bordes de pulso y el espaciamiento de pulso a pulso es constante. En la práctica, el borde anterior se fija por lo general en $t = kT_s$ y sólo el borde posterior se modula, por lo que el espaciamiento de los pulsos es variable. Pero independientemente de cuál o cuáles bordes se modulen, el análisis espectral de la modulación en duración de pulsos es muy complicado. Empleando una aproximación basada en que $\tau_0 \gg T_s$, se puede demostrar que†

$$x_p(t) \approx \frac{A_0 \tau_0}{T_s}[1 + mx(t)] + \text{términos trasladados en frecuencia} \tag{8}$$

y por ello, la modulación en duración de pulsos, como la modulación en amplitud de pulsos, se demodula en forma directa por medio de filtraje de pasabajas, etc.

En la figura 8.12 se ilustra una técnica muy sencilla para la generación de la modulación en duración de pulsos, donde se suman el mensaje y una forma de onda en diente de sierra, para formar la entrada a un *circuito de limitación*. Este circuito, que es en esencia un dispositivo de recorte y cuadrador, produce A_0 volts siempre que la entrada exceda el nivel de limitación y que de otra manera tiene salida cero. La onda resultante es de modulación en duración de pulsos con modulación de borde posterior, con lo que las operaciones de muestreo y modulación se han combinado en un sólo paso. Invirtiendo el diente de sierra se produce modulación de borde anterior; sustituyendo ese diente de sierra por una onda triangular, se produce modulación en ambos bordes. Sin embargo, como puede advertir un lector atento, la duración de pulso final corresponde a las muestras del mensaje ubicadas en el tiempo de borde modulado y no el tiempo de muestreo aparente $t = kT_s$. Así, los valores muestra quedan espaciados de manera no uniforme, y la teoría del muestreo uniforme no es aplicable en forma estricta. Por supuesto que la diferencia entre el muestreo uniforme y el no uniforme es insignificante si la duración del pulso es pequeña en comparación con T_s. Desde el punto de vista de los componentes físicos, se prefiere el muestreo no uniforme a causa de lo simplificado de su instrumentación. Y como regalo, el muestreo no uniforme origina menos distorsión inherente en el mensaje filtrado de salida que el muestreo uniforme.

† El método se bosqueja en el problema 8.17; véase Black (1953, Cap. 17) para detalles.

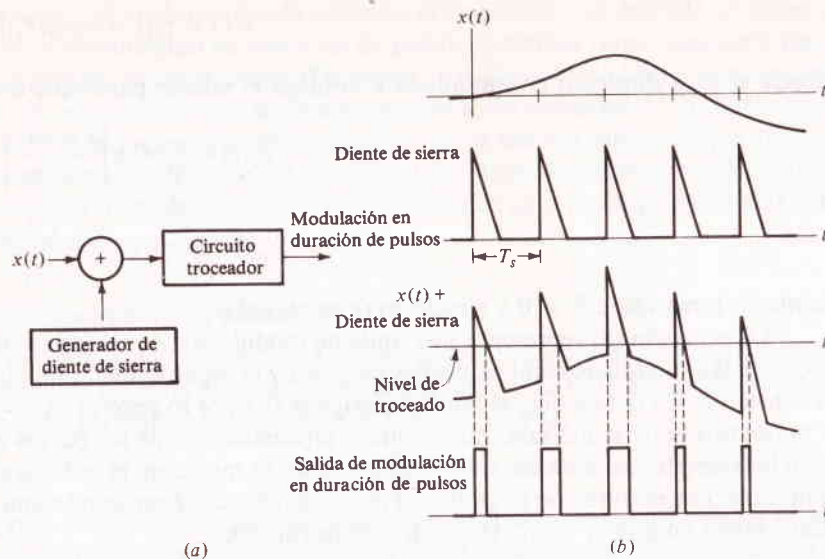



FIGURA 8.12
Generación de la modulación en duración de pulsos. (*a*) Diagrama a bloques; (*b*) formas de onda.

Modulación en posición de pulsos (PPM)

La modulación en duración de pulsos y la modulación en posición de pulsos están íntimamente ligadas, una relación que se hace resaltar en la figura 8.13, donde se muestra cómo se puede generar a la segunda a partir de la primera con modulación de borde posterior. Los pulsos modulados en duración están invertidos y diferenciados, cambiando los bordes modulados en picos positivos modulados en posición. Con un ligero reformado de estos picos se tiene la onda de modulación en posición de pulsos deseada.

Es un hecho que el principal uso de la modulación por duración es la generación (y detección) de la modulación por posición, ya que la modulación en posición de pulsos es superior en una forma marcada respecto a la modulación en duración de pulsos para transmisión de mensajes. Para apreciar el porqué de esto, recuérdese que la información reside en la ubicación en el tiempo de los *bordes* de los pulsos, y no en los pulsos en sí. De esta manera, en forma parecida al caso de la potencia de frecuencia portadora de la AM, la potencia de pulsos de la modulación de pulsos en tiempo es potencia "desperdiciada", y se tendría más eficiencia ¡suprimiendo los pulsos y transmitiendo sólo los bordes! Por supuesto que no se pueden transmitir los bordes sin transmitir los pulsos para definirlos. Pero se pueden enviar pulsos muy cortos para indicar la posición de los bordes, o sea, un proceso equivalente a la modulación en posición de pulsos. La potencia reducida requerida para la modulación en posición de pulsos es una ventaja fundamental sobre la modulación en duración

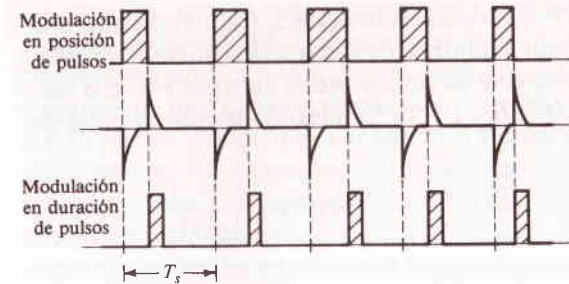



FIGURA 8.13
Generación de la modulación en posición de pulsos a partir de la modulación en duración de pulsos.

de pulsos, una ventaja que será más evidente cuando se examinen las relaciones señal a ruido.

Puesto que la modulación en posición de pulsos con muestreo no uniforme es el tipo más eficiente de modulación de pulsos en tiempo para propósitos de transmisión, se tomará el tiempo para analizar su espectro. El método de análisis merece ser examinado.

Análisis espectral de la modulación en posición de pulsos

Para empezar, sea t_k la ubicación en el tiempo (centro) del pulso k-ésimo. Si el muestreo es uniforme, el pulso k-ésimo conduce al valor muestra en $t = kT_s$, y

$$t_k = kT_s + t_0\,x(kT_s)$$

donde t_0 es la constante de modulación, es decir, el *desplazamiento máximo* con relación a $t = kT_s$. Pero con un muestreo no uniforme, el valor muestra en realidad se extrae en t_k, y no en kT_s, por lo que

$$t_k = kT_s + t_0\,x(t_k) \tag{9}$$

Por definición, la onda de modulación en posición de pulsos es una suma de pulsos modulados en posición de amplitud constante, que se pueden escribir como

$$x_p(t) = \sum_k A_0 p(t - t_k) = A_0[p(t)] * \left[\sum_k \delta(t - t_k)\right]$$

donde A_0 es la amplitud de los pulsos y $p(t)$ su forma. En este punto es posible hacer una simplificación notando que $p(t)$ tendrá (o debe tener) una duración muy pequeña en comparación con T_s. En consecuencia, y para los propósitos que se persiguen, la forma de pulsos puede considerarse como impulsiva, o sea, $p(t) \approx \delta(t)$, por lo que

$$x_p(t) \approx A_0 \sum_k \delta(t - t_k) \tag{10}$$

Si se desea, la ecuación (10) se puede convolucionar con $p(t)$ para expresar la forma no impulsiva.

En sus formas presentes, las ecuaciones (9) y (10) no son adecuadas para una manipulación más detallada; el problema es con el término de posición t_k, el cual no

se puede resolver en forma explícita. Por fortuna, Rowe (1965, Cap. 5) inventó una técnica con la que t se puede eliminar por completo, de la manera siguiente: considérese cualquier función $g(t)$ que tenga un cero de primer orden único en $t = \lambda$; o sea, $g(\lambda) = 0$, $g(t \neq \lambda) \neq 0$, y $g'(\lambda) = dg/dt|_{t=\lambda} \neq 0$. Empleando las propiedades de los impulsos, se puede mostrar que †

$$\delta[g(t)] = \frac{\delta(t - \lambda)}{|g'(t)|}$$

de donde

$$\delta(t - \lambda) = |g'(t)|\,\delta[g(t)] \qquad (11)$$

cuyo segundo miembro es independiente de λ. Por lo tanto, la ecuación (11) se puede emplear para remover a t_k de $\delta(t - t_k)$ si se puede encontrar una función $g(t)$ que satisfaga a $g(t_k)$ y a las otras condiciones, pero que no contenga a t_k.

Supóngase que se toma a $g(t) = t - kT_s - t_0x(t)$, la cual es cero en $t = kT_s + t_0x(t)$. Ahora, para un valor dado de k, hay solo un pulso de modulación en posición de pulsos, que ocurre en $t = kT_s + t_0x(t_k)$. Así, $g(t_k) = t_k - kT_s - t_0x(t_k) = 0$, como se desea. Insertando a $\lambda = t_k$, $g'(t) = 1 - t_0x'(t)$, etc., en la ecuación (11) se tiene

$$\delta(t - t_k) = |1 - t_0\, x'(t)|\,\delta[t - kT_s - t_0\, x(t)]$$

y la onda de modulación en posición de pulsos de la ecuación (10) resulta ser

$$x_p(t) = A_0[1 - t_0\, x'(t)] \sum_k \delta[t - t_0\, x(t) - kT_s]$$

El valor absoluto decae, pues para muchas señales de interés $|t_0x'(t)| < 1$ si $t_0 \ll T_s$. Después se convierte la suma de impulsos a una suma de exponenciales por medio de ‡

$$\sum_{k=-\infty}^{\infty} \delta(t - kT_s) = f_s \sum_{n=-\infty}^{\infty} e^{jn\omega_s t} \qquad (12)$$

para obtener por último

$$x_p(t) = A_0 f_s[1 - t_0\, x'(t)] \sum_{n=-\infty}^{\infty} e^{jn\omega_s[t - t_0 x(t)]}$$

$$= A_0 f_s[1 - t_0\, x'(t)]\left\{1 + \sum_{n=1}^{\infty} 2\cos\,[n\omega_s t - n\omega_s t_0\, x(t)]\right\} \qquad (13)$$

Interpretando la ecuación (13), se ve que la modulación en posición de pulsos con muestreo no uniforme es una combinación de modulación de portadora lineal y

† Véase Friedman (1956).

‡ La ecuación (12) es simplemente el desarrollo en serie exponencial de Fourier de $s_\delta(t)$ puesto que ya se ha demostrado que $c_n = 1/T_s = f_s$. De manera alterna, considérese la transformada inversa de Fourier de $S_\delta(f)$, ecuación (12), (Sec. 8.1).

exponencial, porque cada armónica de f_s está modulada en fase† por el mensaje $x(t)$ y modulada en amplitud por la derivada $x'(t)$. Por lo tanto, el espectro consiste en bandas laterales de AM y de PM centradas en todos los múltiplos de f_s, más una componente de CD y el espectro es un ejercicio tedioso aun para modulación de tono. El primer término de la ecuación (13) sugiere que se puede recuperar el mensaje por medio de filtraje pasabajas e *integración*. Sin embargo, el método de integración no proporciona una ventaja plena sobre las propiedades de reducción de ruido de la modulación en posición de pulsos, por lo que el procedimiento usual es la conversión de modulación en amplitud de pulsos o modulación en duración de pulsos seguida de filtraje pasabajas.

Relaciones señal a ruido

Independientemente de la técnica particular empleada, la demodulación de una onda modulada por pulsos es en forma inherente la reconstrucción del mensaje a partir de los valores muestra. Este punto de vista permite tratar el ruido aditivo en una forma relativamente sencilla. De manera específica, si la señal recibida más ruido es

$$v(t) = K_R\, x_p(t) + n(t)$$

entonces la reconstrucción se puede basar en las muestras contaminadas; es decir, que la salida del convertidor de la figura 8.10 es

$$y_\delta(t) = \sum_k [M_p\, x(kT_s) + \epsilon_k]\delta(t - kT_s) \qquad (14)$$

donde M_p es la constante de modulación que relaciona a $x(kT_s)$ con $x_p(t)$ y ϵ_k es el error introducido por la presencia de $n(t)$. Entonces suponiendo un filtro pasabajas ideal con $B \approx f_s/2$, la salida filtrada es

$$y_D(t) = M_p\, x(t) + \underbrace{\sum_k \epsilon_k \operatorname{senc}(f_s t - k)}_{\xi(t)} \qquad (15)$$

como se desprende de las ecuaciones (14) y (15), (Sec. 8.1) con $K = 1$ y $t_d = 0$ para sencillez. Puesto que el ruido $\xi(t)$ en la ecuación (15) es aditivo, se puede formar la relación señal a ruido en el destino S_D/N_D donde $S_D = M_p^2\overline{x^2}$ y $N_D = \overline{\xi^2}$. En general, $\xi(t)$ no es igual a $\epsilon(t)$ porque la forma de onda de ésta, si existe en forma física, no es necesariamente de banda limitada. Sin embargo, de la ecuación (12), (Sec. 4.5), se tiene en forma inmediata que $\overline{\xi^2} = \overline{\epsilon^2}$ y por ello

$$\left(\frac{S}{N}\right)_D = \frac{M_p^2\overline{x^2}}{\overline{\epsilon^2}} \qquad (16)$$

† Así, el *sistema serrasoide* para la generación de modulación de banda angosta empieza con una onda de modulación en posición de pulsos (véase el Prob. 8.18).

siendo $\overline{\epsilon^2}$ el valor cuadrático medio del término de error.

Se supondrá que los pulsos se transmiten sin modulación de portadora sobre un sistema que tiene un ancho de banda B_T. Entonces $n(t)$ es ruido blanco de banda limitada de densidad de potencia η, y $N_R = \overline{n^2} = \eta B_T$. La potencia promedio de la señal es $S_R = K_R^2 S_T$, como es usual, y la amplitud recibida se denota con $A_R = K_R A_0$.

La modulación en amplitud de pulsos más ruido, se analiza con facilidad, puesto que los valores muestra están contenidos en la amplitud modulada $A_R m x(kT_s)$, y el ruido se agrega en forma directa a la amplitud. En consecuencia, $M_p = A_R m$, $\epsilon_k = n(kT_s)$, $\overline{\epsilon^2} = \overline{n^2} = \eta B_T$, y

$$\left(\frac{S}{N}\right)_D = \frac{A_R^2 m^2 \overline{x^2}}{\eta B_T} = \frac{m^2\overline{x^2}}{1 + m^2\overline{x^2}} \frac{T_s}{\tau} \left(\frac{S}{N}\right)_R$$

o

$$\left(\frac{S}{N}\right)_D = \frac{m^2\overline{x^2}}{1 + m^2\overline{x^2}} \frac{T_s W}{\tau B_T}\gamma \qquad \text{PAM} \qquad (17)$$

donde $\gamma = S_R/\eta W$. Se estableció antes que la modulación en amplitud de pulsos no es mejor que la transmisión directa en banda base en cuanto a ruido se refiere. La ecuación (17) demuestra que esto es cierto, para $T_s \leq 1/2W$ y $B_T \geq 1/2\tau$, por lo que $T_s W/\tau B_T \leq 1$ y $(S/N)_D \leq \gamma/2$. La igualdad se cumple sólo por muestreo en la velocidad de Nyquist con $m^2\overline{x^2} = 1$. Estas condiciones se logran raramente en la práctica; y ni se les ha buscado, dado que el mérito de la modulación en amplitud de pulsos reside en su sencillez en la canalización, y no en su comportamiento ante el ruido.

Volviendo a la modulación en tiempo de pulsos, recuérdese que las muestras del mensaje están contenidas en la posición relativa de los bordes de los pulsos. Si los pulsos recibidos fueran rectangulares en forma completa, el ruido aditivo no tendría efecto sobre los bordes, puesto que el ruido es una perturbación *vertical*. Pero los pulsos rectangulares requieren un ancho de banda de transmisión *infinito*; con ancho de banda finito los pulsos tienen un tiempo de subida diferente de cero $\tau_r \approx 1/2B_T$ y el ruido alteraría la posición del borde como se estudió en la sección 4.4, (Fig. 4.20). Aplicando los resultados de ese estudio,

$$\overline{\epsilon^2} = \overline{\Delta t^2} = \frac{\eta B_T \tau_r^2}{A_R^2} = \frac{\eta}{4 B_T A_R^2}$$

Nótese que el error cuadrático medio *disminuye* conforme *aumenta* el ancho de banda; con lo cual se obtiene la ***reducción de ruido de banda ancha*** prometida.

Como en los parámetros de la señal, la variación en la duración de los pulsos o en la posición de ellos, es $t_0 x(kT_s)$, haciendo $t_0 = m\tau_0$ en la modulación en duración

de pulsos, por lo que $M_p = t_0$. La potencia promedio es $S_R = (\tau/T_s)A_R^2 = \tau f_s A_R^2$, puesto que la amplitud de los pulsos es constante. Por

$$\left(\frac{S}{N}\right)_D = \frac{t_0^2\overline{x^2} 4 B_T S_R}{\tau\eta f_s} = \frac{4 t_0^2 B_T W}{\tau f_s}\overline{x^2}\gamma \qquad (18)$$

lo cual se aplica tanto a la modulación en duración de pulsos como a la modulación en posición de pulsos sin modulación de portadora. En el caso de la primera, se debe interpretar a τ como la duración de pulso *promedio* τ_0.

La única diferencia entre uno y otro tipo de modulación, como lo refleja la ecuación (18) es la duración más pequeña de pulso en la modulación en posición de pulsos. Esta diferencia se puede hacer más evidente considerando lo mejor que se puede hacer para un ancho de banda dado B_T. Es claro que el aumentar al máximo a $(S/N)_D$ requiere de la mínima frecuencia de muestreo ($f_s = 2W$) con t_0 máxima y duración promedio del pulso mínima. En la modulación en posición de pulsos, t_0 y τ son independientes y se pueden tratar en forma separada. La mínima duración de pulso posible es $\tau \approx 2\tau_r \approx 1/B_T$, para la cual el pulso degenera en una forma triangular. Se debe limitar el máximo desplazamiento del pulso de tal manera que el *intervalo* de las posibles posiciones de pulso, $kT_s \pm 2t_0$, no exceda el tiempo entre muestras, o sea, $t_0 \leq T_s/2 = 1/4W$. Considerando que todos los valores son los óptimos, se tiene

$$\left(\frac{S}{N}\right)_{D_{\text{máx}}} = \frac{1}{8}\left(\frac{B_T}{W}\right)^2 \overline{x^2}\gamma \qquad \text{Modulación en posición de pulsos} \qquad (19)$$

el cual muestra el intercambio de ancho de banda por relación señal a ruido. Sin embargo, a los sistemas prácticos puede faltarles para alcanzar este máximo un orden de magnitud, o más.

Una optimización similar aplicada a la modulación en duración de pulsos resulta en $m\tau_0 \approx 1/4W$, por lo que

$$\left(\frac{S}{N}\right)_{D_{\text{máx}}} = \frac{1}{2}\frac{B_T}{W}\overline{x^2}\gamma \qquad \text{PDM} \qquad (20)$$

Nótese que la relación señal a ruido se incrementa en forma lineal con la relación de anchos de banda B_T/W en vez de con $(B_T/W)^2$. Además, para obtener la ecuación (20) la duración promedio del pulso debe ser la mitad del tiempo entre las muestras. Bajo esta condición, el ciclo de trabajo es $\tau_0/T_s = ½$, con lo que se pierden las ventajas prácticas de la modulación de pulsos con un ciclo de trabajo pequeño.

Es de interés particular comparar estos resultados con los de la modulación exponencial. Recuérdese, por ejemplo, que la FM de banda ancha sin desacentuación produjo

$$\left(\frac{S}{N}\right)_D = \frac{3}{4}\left(\frac{B_T}{W}\right)^2 \overline{x^2}\gamma$$

ancho de banda, suponiendo a $B_T = W \gg 1$. La modulación en posición de pulsos es mucho mejor si damos una relación señal a ruido de aproximadamente 8 dB por debajo en relación con la FM. Aparece así que, mientras la modulación de pulsos en tiempo sí proporciona una reducción de ruido de banda ancha, la reducción es mucho menos efectiva que en la modulación exponencial de portadora. Pero se debe recordar que la potencia promedio S_T se entrega como pulsos de alta potencia y de corta duración, mientras que a S_T se le debe generar en forma continua en FM. Las limitaciones prácticas del suministro de portadora pueden favorecer la operación pulsada.

EJERCICIO 8.4 Supóngase que una señal modulada por pulsos se transmite como pulsos de RF con potencia promedio S_{RF} y ancho de banda B_{RF}. Demuéstrese que las ecuaciones de la (17) a la (20) se deben modificar por medio de las sustituciones

$$B_T \to \frac{B_{\text{RF}}}{2} \qquad S_R \to 2S_{\text{RF}} \tag{21}$$

Demuéstrese después que la modulación en posición de pulsos con modulación de portadora cae a por lo menos 11 dB debajo de la FM de banda ancha.

Nivel de umbral

Para concluir este análisis, se debe mencionar que la relación de anchos de banda no se puede incrementar sin límite, dado que la modulación de pulsos en tiempo experimenta un *efecto de umbral* en forma similar a la FM. Es cierto que $\overline{\epsilon^2} \to 0$ conforme $B_T \to \infty$. Pero conforme $N_R = \eta B_T$ aumenta, el voltaje rms de ruido se incrementa a su vez, con lo que los picos ocasionales de ruido serán tomados erróneamente por pulsos de la señal. Si tales *pulsos falsos* ocurren muy a menudo, el mensaje deseado se perderá por completo, porque la señal reconstruida no tendrá ninguna relación con $x(t)$.

Se dirá que los pulsos falsos no ocurren con la suficiente frecuencia si $p(n > A_R) \leq 0.01$. Suponiendo ruido gaussiano, la correspondiente condición de umbral es aproximadamente

$$A_R \geq 2\sqrt{N_R}$$

es decir, que la amplitud del pulso debe ser bastante intensa para "elevar" el ruido por lo menos dos veces su valor rms. (Ésta es la misma condición que existe en la *sensibilidad tangencial* de los sistemas de radares pulsados). Así, puesto que $A_R^2 = S_R/\tau f_s$ en las modulaciones en duración y en posición de pulsos,

$$\left(\frac{S}{N}\right)_{R\text{-ésima}} = 4\tau f_s \tag{22a}$$

y

$$\gamma_{\text{ésima}} = \frac{B_T}{W} 4\tau f_s \geq 8 \tag{22b}$$

por lo que el nivel de umbral es menor en forma apreciable que en el caso de la modulación exponencial.

8.3 MODULACIÓN DE PULSOS CODIFICADOS: PCM, DM Y DPCM

Todos los tipos de modulación considerados hasta aquí, sean de pulsos o de CW, han sido representaciones analógicas del mensaje. La modulación por pulsos codificados (PCM) es completamente diferente en concepto; es una modulación digital en la que el mensaje se representa por medio de un grupo codificado de pulsos digitales (amplitudes discretas). La modulación en delta (DM) y la modulación diferencial por pulsos codificados (DPCM) son variantes de la modulación por pulsos codificados. El razonamiento en que se apoya el procedimiento de digitalización es como sigue.

En la modulación analógica, el parámetro modulado varía en forma continua y puede tomar *cualquier* valor de los correspondientes al intervalo del mensaje. Cuando la onda modulada se altera con ruido, no existe en el receptor forma de distinguir el valor transmitido exacto. Supóngase, sin embargo, que se permiten sólo unos pocos valores *discretos* para el parámetro modulado; si la separación entre estos valores es grande en comparación con las perturbaciones de ruido, será algo sencillo decidir en el receptor con precisión los valores específicos que fueron enviados. Así, se pueden eliminar de manera virtual los efectos del ruido aleatorio, lo cual constituye el objetivo de la modulación por pulsos codificados. En forma colateral, la propiedad de amplitudes discretas se puede emplear en los sistemas de este tipo de modulación para larga distancia con repetidores *regenerativos* descritos en la sección 4.5, con lo que se gana una ventaja adicional sobre las demás formas de transmisión analógica.

Pero surge ahora la pregunta: ¿Cómo se representa un mensaje analógico en forma digital? La respuesta cae dentro del terreno del muestreo, de la cuantificación y de la codificación.

Cuantificación y codificación

En la figura 8.14 se muestra el diagrama con los elementos de la generación de modulación por pulsos codificados. La señal continua primero se filtra a pasabajas (¿por qué?) y después se muestrea para dar $x_s(t)$. Los valores muestra se redondean o cuantifican al valor discreto predeterminado más próximo o *nivel cuántico*. La señal muestreada y cuantificada resultante $x_{sq}(t)$ es discreta en tiempo (en virtud del muestreo) y en amplitud (en virtud de la cuantificación). Por último, $x_{sq}(t)$ se opera mediante un codificador que convierte las muestras cuantificadas en *palabras de código digital* apropiadas, con una palabra de código por cada muestra, y genera la correspondiente señal de banda base de modulación por pulsos codificados como

FIGURA 8.14
Sistema para la generación de la modulación por pulsos codificados.

una forma de onda digital. Nótese, que vista desde otro punto de vista, la figura 8.14 es un convertidor analógico a digital (A/D).

De manera obvia, los parámetros de la señal codificada dependen del número de niveles cuánticos Q, porque cada palabra de código debe representar exclusivamente una de las posibles muestras cuantificadas. Para indagar la relación, sea ν el número de dígitos en la palabra de código, teniendo cada una uno de los valores discretos. Puesto que hay μ^ν palabras de código posibles diferentes, se requiere que $\mu^\nu \geq Q$ para una codificación única. Por lo tanto, cuando se eligen los parámetros de tal manera que la igualdad se cumpla,

$$\mu^\nu = Q \qquad \nu = \log_\mu Q \tag{1}$$

Examinando la ecuación (1), se ve que si $\mu = Q$, entonces $\nu = 1$, y la señal cuantificada no requiere de traslación de código. Sin embargo, en general, $\mu < Q$ y $\nu > 1$. La forma más común de modulación por pulsos codificados en la *binaria*, para la cual $\mu = 2$; el número de niveles cuánticos se toma entonces como alguna potencia de 2, a saber, $Q = 2^\nu$.

La figura 8.15 ilustra estas operaciones para modulación por pulsos codificados binaria. Se muestran ocho niveles cuánticos, correspondientes a $\pm$ 1/8, $\pm$ 3/8, . . ., $\pm$ 7/8, por lo que los niveles están espaciados de manera uniforme en $2 \times 1/8 = 2/Q$ e incluyendo el redondeo, ellos cubren el intervalo de $-$ 1.0 a + 1.0 lo cual es consistente con la convención de normalización $|x(t)| \leq 1$. Se asigna un número de código 0, 1, . . .,7 a cada nivel cuántico y la palabra de código binario es simplemente el equivalente binario del número de código, por ejemplo, 5 → **101**, etc. Se necesitan tres dígitos binarios por palabra de código, puesto que $\nu = \log_2 8 = \log_2 2^3 = 3$.

Como se requieren varios dígitos para cada muestra de mensaje, resulta que el ancho de banda de la modulación por pulsos codificados será mucho mayor que el ancho de banda del mensaje. Una estimación del ancho de banda se obtiene como sigue. Las muestras cuantificadas ocurren a una velocidad de $f_s \geq 2W$ muestras por segundo, por lo que debe haber $r = \nu f_s$ dígitos por segundo. Recordando la relación ancho de banda velocidad de señalización de la sección 4.5, se tiene que

$$B_{T_{BB}} \geq \frac{r}{2} = \frac{\nu f_s}{2} \geq \nu W \tag{2}$$

El ancho de banda de la modulación por pulsos codificados en banda base es así un mínimo de $\nu = \log_\mu Q$ veces el ancho de banda del mensaje.

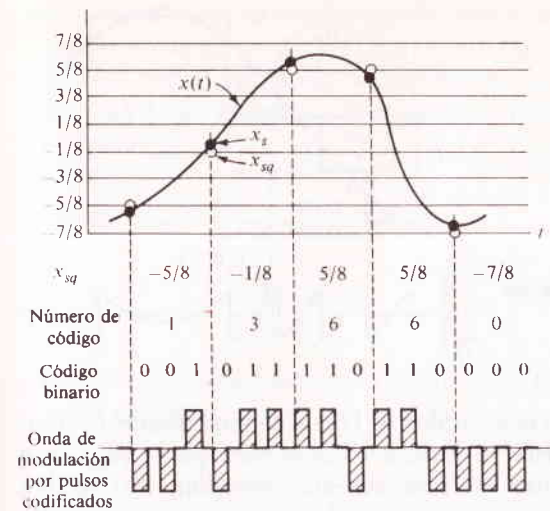



FIGURA 8.15

Como paso final de la generación de la modulación por pulsos codificados, la señal de banda base puede modular una portadora de RF para propósitos de transmisión. La modulación de la portadora puede ser de *manipulación por corrimiento de amplitud* (ASK), *manipulación por corrimiento de fase* (PSK) o *manipulación por corrimiento de frecuencia* (FSK). (Estas son formas de modulación de CW para señales digitales analizadas en el capítulo 10.) Por supuesto que la modulación de portadora resulta en un ancho de banda de transmisión aún mayor que el de la ecuación (2) pero sólo se prestará atención al caso de banda base.

Ruido de cuantificación

La figura 8.16 muestra la porción de un receptor para la modulación por pulsos codificados que sigue a la demodulación de portadora, si la hay. La forma de onda de la modulación por pulsos codificados analógica contaminada por ruido aleatorio $n(t)$ se opera por medio de un convertidor A/D que regenera las palabras de código digitales (más errores) como se describió en la sección 4.5. De estas palabras de código, el decodificador determina los valores muestra cuantificados (de nuevo con errores) y genera a $x_{sq}(t)$ la cual se procesa por medio de un filtro pasabajas para obtener la señal analógica de salida.

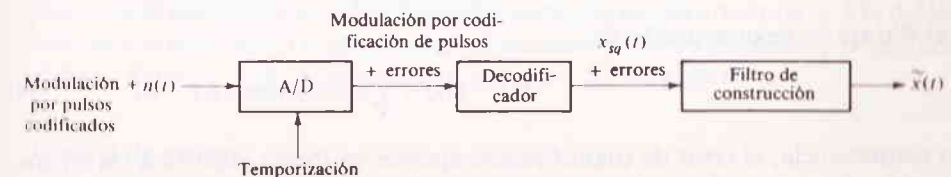



FIGURA 8.16
Receptor para modulación por pulsos codificados.

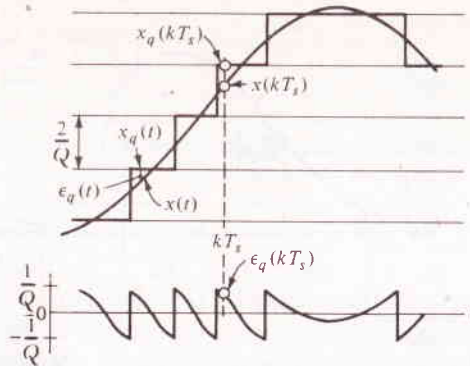



FIGURA 8.17
Error de cuantificación en la modulación por codificación de pulsos.

Si la relación señal a ruido en el convertidor A/D es sólo modestamente grande, la probabilidad de error es pequeña, lo suficiente como para que se haga caso omiso de los efectos del ruido aleatorio. No obstante esta condición, $\tilde{x}(t)$ no será idéntica al mensaje $x(t)$ puesto que la entrada del filtro pasabajas es $x_{sq}(t)$ y no $x_s(t)$; es decir, que la reconstrucción se basa en las muestras cuantificadas más que en los valores muestra exactos. Además, no hay forma de obtener valores exactos en el receptor; esa información se descartó en el transmisor durante el proceso de cuantificación. Por lo tanto, es imposible una reconstrucción perfecta del mensaje en los sistemas de codificación de pulsos, aún cuando el ruido aleatorio tenga una influencia despreciable.

Como resultado, el efecto de cuantificación es una limitación básica de los sistemas codificados, como el ruido aleatorio es una limitación de los sistemas analógicos convencionales. Es este efecto de cuantificación él que se desea examinar más detalladamente.

Para efectos de análisis, se descompone a $x(t)$ en una señal cuantificada en forma de escalón $x_q(t)$ en una señal cuantificada en forma de escalón $x_q(t)$ y en un *término de error de cuantificación* $\epsilon_q(t)$, tal que

$$x_q(t) = x(t) + \epsilon_q(t) \qquad (3a)$$

como se ilustra en la figura 8.17. Dado que es intercambiable el orden de muestreo y cuantificación, los valores muestra cuantificados conforme se decodifican en el receptor se pueden escribir en la forma

$$x_q(kT_s) = x(kT_s) + \epsilon_q(kT_s) \qquad (3b)$$

y el filtraje de reconstrucción da

$$\tilde{x}(t) = x(t) + \sum_k \epsilon_q(kT_s)\,\mathrm{senc}(f_s t - k) \qquad (4)$$

En consecuencia, el error de cuantificación aparece en forma idéntica a los errores de ruido aleatorio en la modulación analógica de pulsos, ecuación (15), (Sec. 8.2). Por consiguiente, se designa a $\overline{\epsilon_q^2}$ como el ruido de cuantificación.

Este ruido se evalúa con facilidad cuando los niveles cuánticos tienen espaciamiento *uniforme* $2/Q$ así $|\epsilon_q(t)| \le 1/Q$ —(véase la figura 8.17). A falta de información por el contrario, se supone que ϵ_q está uniformemente distribuido sobre $[-1/Q, 1/Q]$; es decir, que su función de densidad de probabilidad es $p(\epsilon_q) = Q/2$, $|\epsilon_q| \le 1/Q$. Así

$$\overline{\epsilon_q^2} = \int_{-1/Q}^{1/Q} \epsilon_q^2 \frac{Q}{2}\, d\epsilon_q = \frac{1}{3Q^2} \qquad (5)$$

y

$$\left(\frac{S}{N}\right)_D = \overline{x^2}/\overline{\epsilon_q^2} = 3Q^2\overline{x^2} \qquad (6)$$

la cual es independiente de la modulación de portadora, potencia transmitida y del ruido aleatorio, puesto que aquí no se hace caso de los errores de decodificación.

Es obvio que la calidad en el comportamiento se incrementa con Q, el número de niveles cuánticos. Esto simplemente reitera que si se están empleando muchos niveles cuánticos de espaciamiento pequeño, las muestras cuantificadas se aproximan mucho a los valores muestra exactos y la salida se puede hacer tan cercana a $x(t)$ como se desee. Sorprende que pocos niveles produzcan un comportamiento satisfactorio; por ejemplo, $Q = 2^7$ o 2^8 es normal en telefonía. Sin embargo, se debe señalar que el error rms se fija en $1/\sqrt{3}\,Q$ sin tomar en cuenta el valor instantáneo de $x(t)$. En consecuencia, si $|x(t)|$ es pequeña en periodos prolongados, la relación señal a ruido aparente será mucho menor que el valor de diseño. El efecto es agudo particularmente si la forma de onda del mensaje tiene un *factor de cresta* grande (la relación de amplitud pico al valor rms), porque entonces $|x(t)| \ll 1$ la mayor parte del tiempo, y $\overline{x^2} \ll 1$.

Para la transmisión de señales de audio, que se caracterizan de manera típica por tener factores de cresta grandes, es ventajoso reducir el espaciamiento entre los niveles cuánticos, con un espaciamiento cercano a cero y otro grande en los extremos. Puede resultar en $\epsilon_q(t)$ una cuantificación *no unifome* adecuada y en forma sustancial proporcional a $|x(t)|$, enmascarándose así el ruido con la señal en cuanto al oyente concierne. En la práctica, la cuantificación ahusada se afecta con niveles espaciados de manera uniforme, *comprimiéndose* no linealmente el mensaje antes del muestreo; un expansor complementario restituye la forma de onda en el receptor.

EJERCICIO 8.5 El error de cuantificación para un sistema de modulación por pulsos codificados con $\mu = 2$ se especifica para que no sea mayor de $\pm$ 5% del intervalo pico a pico de $x(t)$. Encuéntrese el número mínimo de dígitos por palabra de código. *Resp.*: $\nu = 4 > \log_2$, puesto que debe ser un entero.

Umbral de error

Ahora considérese la situación en la cual no se pueden despreciar los errores de decodificación causados por el ruido aleatorio, por lo que se tiene *ruido de decodifica-*

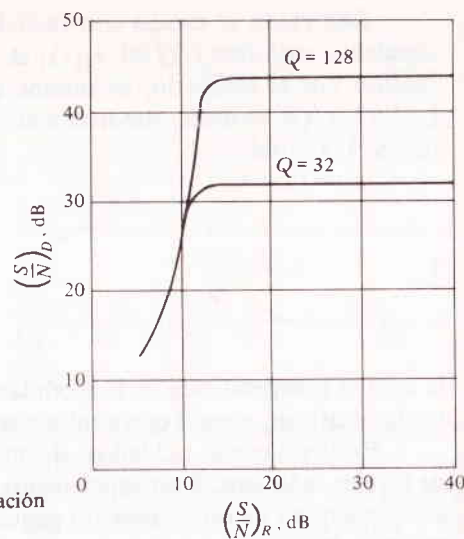



FIGURA 8.18
Comportamiento bajo ruido de la modulación por codificación de pulsos.

ción y de cuantificación. El cálculo del ruido de decodificación cuadrático medio $\overline{\epsilon_d^2}$ es complicado porque el número de errores por palabra de código es aleatorio y la severidad de un error depende de donde ocurra, siendo más serio en el digito principal (más significativo) que otros errores en la misma palabra. Suponiendo que la probabilidad de error por dígito P_e no es muy grande, se puede demostrar para la modulación binaria por pulsos codificados que †

$$\overline{\epsilon_d^2} = \frac{4(Q^2 - 1)P_e}{3Q^2} \tag{7a}$$

y

$$\left(\frac{S}{N}\right)_D = \frac{\overline{x^2}}{\overline{\epsilon_q^2} + \overline{\epsilon_d^2}} = \frac{3Q^2\overline{x^2}}{1 + 4(Q^2 - 1)P_e} \tag{7b}$$

donde P_e depende de $(S/N)_R$, la relación señal a ruido en el receptor.

En la figura 8.18 se construye la gráfica de $(S/N)_D$ conta $(S/N)_R$, (ambas en decibeles) para modulación binaria por pulsos codificados con $Q = 32$ y 128, suponiendo ruido gaussiano y $\overline{x^2} = ½$. El descenso abrupto a baja $(S/N)_R$ sugiere en forma contundente el efecto de umbral de FM donde, abajo del umbral, se comprueba que el mensaje queda mutilado al grado de ser inútil. En la modulación por pulsos codificados, también se presenta el caso de la mutilación puesto que se aprecia con facilidad que la forma de onda reconstruida tiene poca semejanza con el mensaje original cuando los errores de decodificación son frecuentes. En otras palabras, el mensaje resulta mutilado por el ruido de decodificación.

† Downing (1964, cap. 7)

Para establecer el nivel de umbral en la modulación por pulsos codificados, se expresará que los errores de decodificación son despreciables si $P_e \leq 10^{-4}$. Luego, tomando a P_e como se dio en la ecuación (15), (Sec. 4.5), es decir;

$$P_e = 2\left(1 - \frac{1}{\mu}\right)Q\left[\sqrt{\frac{3}{\mu^2 - 1}\left(\frac{S}{N}\right)_R}\right] \leq 10^{-4} \tag{8}$$

se resuelve para $(S/N)_{R\text{-ésima}}$ para obtener

$$\left(\frac{S}{N}\right)_{R\text{ésima}} \approx 5(\mu^2 - 1) \tag{9}$$

Por último, puesto que $\gamma = (B_T/W)(S/N)_R$ y $B_T \geq \nu W$,

$$\gamma_{\text{ésima}} = 5\frac{B_T}{W}(\mu^2 - 1) \geq 5\nu(\mu^2 - 1) \tag{10}$$

Una implicación sutil pero importante de la ecuación (9) es su limitación en μ, el número de dígitos diferentes. Específicamente y para cualquier valor de $(S/N)_R$ el sistema está arriba del umbral sólo si

$$\mu^2 \leq \frac{1}{5}\left(\frac{S}{N}\right)_R + 1 \tag{11}$$

lo cual implica una limitación mayor en Q puesto que $\mu^2 = Q^{2/\nu}$.

Modulación por pulsos codificados contra modulación analógica

Arriba del umbral, la relación señal a ruido en la modulación por pulsos codificados es constante y fija en el transmisor por el número de niveles cuánticos. Falta demostrar que la modulación por pulsos codificados tiene las características deseables de reducción de ruido de banda ancha con las características indeseables de mutilación abajo del umbral. Para este propósito, se supone que la frecuencia de muestreo es próxima a la velocidad de Nyquist por lo que el ancho de banda base es $B_T \approx \nu W$, por la ecuación (2). Entonces $Q = \mu^\nu \approx \mu^2$ donde $\mathscr{B} = B_T/W$ es la relación de anchos de banda, y la ecuación (6) resulta ser

$$\left(\frac{S}{N}\right)_D \approx 3\mu^{2\mathscr{B}}\overline{x^2} \tag{12}$$

lo cual demuestra que la reducción de ruido es un intercambio *exponencial* de ancho de banda por la relación señal a ruido. Este intercambio tiene más consecuencias que aquel de la modulación analógica de banda ancha, en donde $(S/N)_D$ se incrementa en forma lineal o conforme al cuadrado de la relación de anchos de banda.

La figura 8.19 muestra el comportamiento de la modulación binaria por pulsos codificados ($\mu = 2$) según se describió por medio de las ecuaciones (10) y (12), para tres valores de la relación de anchos de banda. También muestra, para compa-

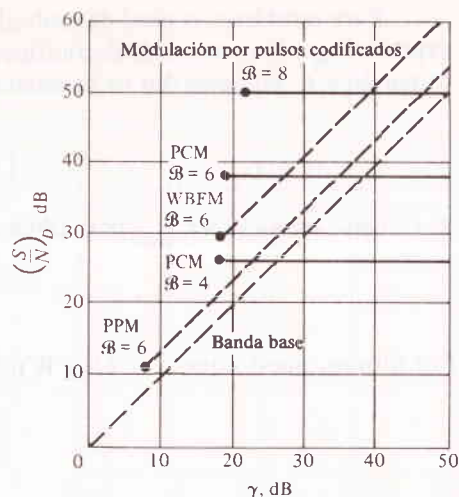



FIGURA 8.19

ración, el comportamiento correspondiente de varios sistemas analógicos. Las curvas para la modulación por pulsos codificados y la modulación en posición de pulsos son para la transmisión directa sin modulación de portadora, y en todos los casos se supone que $\overline{x^2} = ½$. Como es usual, los puntos de umbral se indican por medio de puntos gruesos.

El examen de la figura revela que, en aras de la eficiencia, los sistemas de modulación por pulsos codificados se deben operar precisamente sobre el umbral, dado que el incremento de la potencia transmitida más allá del valor $\gamma_{\text{-ésima}}$ no ocasiona mejoría en $(S/N)_D$; lo último queda determinado sólo por el número de niveles cuánticos o, en forma equivalente, la relación de anchos de banda. Pero incluso en el umbral, la modulación por pulsos codificados no demuestra ser tan buena como se podría haber esperado. Además, la instrumentación es en forma considerable más compleja y costosa que la de los sistemas no codificados. (Precisamente por esta razón, a la modulación por pulsos codificados se le declaró completamente impráctica antes de que surgiera la electrónica digital de estado sólido de alta velocidad en las postrimerías de la década de los cincuentas, unos 20 años después de la invención de la modulación por pulsos codificados.) ¿Por qué entonces este tipo de modulación tiene tanta demanda? Ésta es una intrincada y controvertida interrogante con varias respuestas.

1 Para una relación de anchos de banda dada, la modulación por pulsos codificados cerca del umbral resulta superior a todas las otras formas de modulación *de pulsos*, un factor significativo si se desea la multicanalización por división en tiempo.

2 La modulación por pulsos codificados tiene un margen pequeño, pero definido, sobre la FM a bajas relaciones señal a ruido. E incluso una reducción de potencia de 3 dB, con un factor de 2 puede significar la diferencia entre el éxito o el fracaso en aplicaciones de potencia mínima.

3 Un sistema de modulación por pulsos codificados planeado para transmisión analógica de mensajes se adapta con facilidad a otras señales de entrada, en particular datos digitales, elevando con ello la flexibilidad e incrementando la utilización del sistema .†

4 En virtud de la capacidad de *regeneración*, la modulación por pulsos codificados es precisamente ventajosa para sistemas que tienen muchas estaciones repetidoras. Además, con respecto a la telefonía de larga distancia, a esto se le ha considerado como el verdadero desquite de la modulación por pulsos.

Por lo tanto, a la modulación por pulsos codificados se le debe dar la debida consideración para aplicaciones en que se involucre la multicanalización por división en tiempo, potencia mínima, una diversidad de tipos de mensajes (es decir, analógicos y digitales), o muchas estaciones repetidoras. Dado que muchos de estos factores están en la transmisión telefónica de larga distancia, la modulación por pulsos codificados se presenta como el modo del futuro en telefonía.‡ Sin embargo, en muchas aplicaciones de rutina, el costo de los componentes físicos para modulación codificada resulta casi siempre prohibitivo en comparación con los correspondientes a la modulación analógica —una conclusión a revocarse dados los avances en los circuitos integrados.

EJERCICIO 8.6 A partir de las ecuaciones (10) y (12), demuéstrese que un sistema de modulación por pulsos codificados en el punto de umbral tiene precisamente

$$\left(\frac{S}{N}\right)_{D\,\text{ésima}} = 3\left[1 + \left(\frac{\gamma_{\text{ésima}}}{5\mathscr{B}}\right)\right]^{\mathscr{B}} \overline{x^2} \tag{13}$$

Compárese esto con la FM de banda ancha haciendo $\Delta = \mathscr{B}/2 \gg 1$ en la ecuación (24), (Sec. 7.5).

Modulación en delta ★

La modulación en delta es *descendiente directa* de la modulación por pulsos codificados y tiene la ventaja de que sus componentes físicos están muy simplificados; además, es el método conocido más simple para la conversión de una señal analógica a la forma digital. Como contraste con estas ventajas, la modulación en delta requiere un ancho de banda mayor que la modulación por pulsos codificados. Sin embargo, para señales de voz, los refinamientos recientes han reducido los requisitos de ancho de banda hasta el punto de que la modulación en delta se ha vuelto un fuerte competidor de la modulación por pulsos codificados. Aquí se bosquejarán los principios operacionales de la modulación en delta, dejando los detalles para que sean revisados por los lectores interesados.

† En un interesante articulo de estudio, Franklin y Law (1966) examinan los problemas de la interconexión y compatibilidad en la modulación por pulsos codificados.

‡ Véase Reeves (1965) para una especulación adicional, con estas líneas del pionero de la modulación por pulsos codificados.

‡ Un buen punto de partida es Schindler (1970).

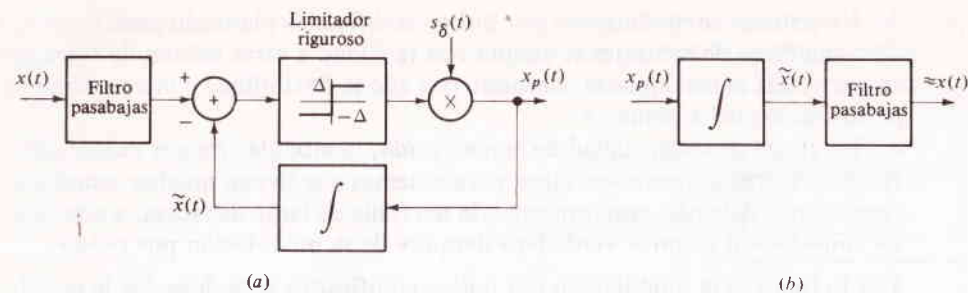



FIGURA 8.20
Modulación en delta. (*a*) Modulador; (*b*) demodulador.

Contrastado esto con la ecuación (20), muestra que la modulación en duración de pulsos queda lejos de alcanzar la reducción de ruido de la FM que tenga el mismo

La figura 8.20*a* es el diagrama funcional a bloques de un modulador en delta. El mensaje $x(t)$ se compara con una aproximación al escalón $\tilde{x}(t)$ por sustracción, pasándose la diferencia a través de un limitador riguroso cuya salida es igual a $\pm\Delta$ dependiendo del signo de $x(t) - \tilde{x}(t)$. Esta, a su vez, modula a la onda de muestreo ideal $s_\delta(t)$ para producir

$$x_p(t) = \sum_k \Delta \operatorname{sgn}[x(kT_s) - \tilde{x}(kT_s)]\, \delta(t - kT_s) \tag{14}$$

una forma de onda de impulso de la cual se genera $\tilde{x}(t)$ por integración. Puesto que hay sólo dos pasos de impulso posibles en $x_p(t)$, la señal que en realidad se transmite es una forma de onda binaria. El demodulador (Fig. 8.20*b*) está formado de un integrador y un filtro pasabajas que originan $x(t)$ más ruido de cuantificación.

Para hacer más claras estas operaciones, la figura 8.21 muestra las formas de onda típicas $x(t)$, $\tilde{x}(t)$ y $x_p(t)$. En primer lugar, sea $\tilde{x}(t) < x(t)$ por lo que el primer impulso tiene $+\Delta$ de ponderación. Cuando se realimenta e integra, ese impulso produce un cambio a escalón en $\tilde{x}(t)$ de altura $+\Delta$. Este proceso continúa durante el intervalo de elevación hasta que $\tilde{x}(t)$ excede a $x(t)$ y da lugar a un impulso negativo. Si entonces $x(t)$ permanece constante, $\tilde{x}(t)$ presenta un comportamiento de exploración conocido como *ruido inactivo*. Cuando $x(t)$ está cambiando, $\tilde{x}(t)$ le sigue en

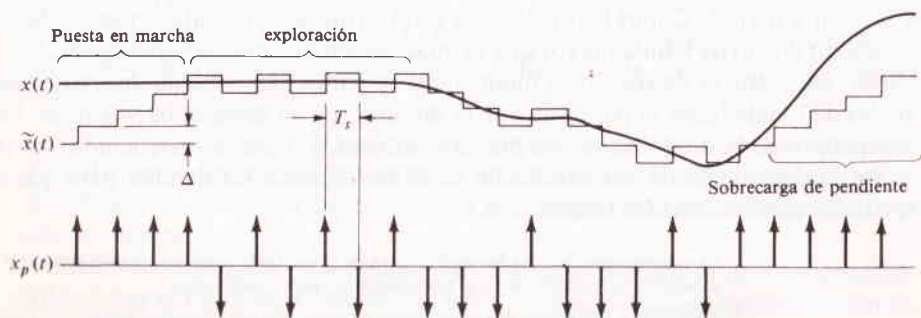



FIGURA 8.21
Formas de onda de la modulación en delta.

forma de escalón a menos que la velocidad de cambio sea muy grande, como se ilustra a la derecha de la figura. Este fenómeno de *sobrecarga de pendiente* es una limitación básica de la modulación en delta.

Aparte de la sobrecarga de pendiente, $\tilde{x}(t)$ se aproxima en forma razonable a $x(t)$ —sobre todo si Δ y T_s son pequeñas— y el filtraje de pasabajas en el demodulador mejora más la aproximación. Pero obsérvese que $x(t)$ no es la señal transmitida. En su lugar, la señal transmitida es una representación binaria de $x_p(t)$ y los dígitos binarios indican sólo la *polaridad de la diferencia* entre $x(t)$ y $\tilde{x}(t)$ en $t = kT_s$; de ahí el nombre de modulación en delta.

Para analizar el comportamiento de la modulación en delta, primero se deduce una condición para prevenir la sobrecarga de pendiente con modulación de tono, $x(t) = A_m \cos 2\pi f_m t$. La máxima pendiente del mensaje es entonces

$$\left[\frac{dx(t)}{dt}\right]_{\text{máx}} = 2\pi f_m A_m \leq 2\pi W$$

donde el límite superior proviene de las convenciones de mensaje $A_m \leq 1$ y $f_m \leq W$. Ahora la pendiente máxima de $\tilde{x}(t)$ es $\Delta / T_s = \Delta f_s$, por lo que una condición suficiente para que no haya sobrecarga de pendiente es

$$f_s \geq \frac{2\pi W}{\Delta} \tag{15}$$

y por lo tanto $f_s \gg 2W$ si $\Delta \ll 1$, requiriéndose lo último para que $\tilde{x}(t)$ sea una buena aproximación de $x(t)$. La ecuación (15) es muy conservadora a menos que el espectro del mensaje sea plano sobre W. De manera más típica, los espectro de mensaje desaparecen muy abajo de $f = W$ y puede relajarse la condición. Así, si existen algunas frecuencias $f_0 < W$ tales que

$$G_x(f) \leq \left(\frac{f_0}{f}\right)^2 G_x(f_0) \qquad f_0 \leq |f| \leq W \tag{16}$$

entonces $dx(t)dt \leq 2\pi f_0$ y, en lugar de la ecuación (15),

$$f_s \geq \frac{2\pi f_0}{\Delta} \tag{17}$$

Las señales de voz por lo general satisfacen la ecuación (16) con $f_0 \approx 800$ Hz en comparación con $W \approx 4$ kHz, por lo que la ecuación (17) representa una reducción significativa de la frecuencia de muestreo y del ancho de banda de transmisión.

Volviendo al ruido de cuantificación, se escribe $\tilde{x}(t) = x(t) + \epsilon(t)$ en donde, de la figura 8.21, $|\epsilon(t)| = |\tilde{x}(t) - x(t)| \leq \Delta$ en ausencia de sobrecarga de pendiente. Si suponemos como antes que $\epsilon(t)$ tiene una distribución uniforme, el error cuadrático medio es igual a $\overline{\epsilon^2} = \Delta^2/3$. No se puede, sin embargo, tomar a $\overline{\epsilon^2}$ como el ruido de cuantificación de salida N_D porque el filtro pasabajas de la figura 8.20*b* opera sobre la señal de escalón $\tilde{x}(t)$, más aún que sobre los valores muestra cuantificados de reconstrucción de la modulación por pulsos codificados. Para encontrar a N_D, se

hace la suposición razonable de que $G_\epsilon(f)$ es esencialmente constante sobre $|f| \leq W$ y $G_\epsilon(0) \approx \overline{\epsilon^2} T_s = \overline{\epsilon^2}/f_s$ (¿por qué?). Por lo tanto,

$$N_D = \int_{-W}^{W} G_\epsilon(f)\, df \approx \frac{W\overline{\epsilon^2}}{f_s} = \Delta^2 \frac{W}{3f_s}$$

y

$$\left(\frac{S}{N}\right)_D = \left(\frac{3f_s}{\Delta^2 W}\right)\overline{x^2}$$

Por último, empleando la ecuación (17) para eliminar a Δ e insertando la relación de anchos de banda $\mathscr{B} = B_T/W \geq f_s/2W$, se tiene

$$\left(\frac{S}{N}\right)_{D\,\text{máx}} = \frac{6}{\pi^2}\left(\frac{W}{f_0}\right)^2 \mathscr{B}^3\overline{x^2} \tag{18}$$

lo cual también incluye el caso de la ecuación (15) haciendo $f_0 = W$.

Si tomamos como base la ecuación (18) se puede decir que el comportamiento de la modulación en delta cae un poco entre la modulación por pulsos codificados y la modulación en posición de pulsos, puesto que la reducción de ruido de banda ancha cambia con $\mathscr{B}^3$. Como en la modulación por pulsos codificados, la señal transmitida es digital, por lo que se permiten los repetidores regenerativos, mientras que el equipo terminal es mucho menos complejo que en aquélla. Varias modificaciones en la modulación en delta prometen reducir los problemas de sobrecarga de pendiente y de ruido inactivo.

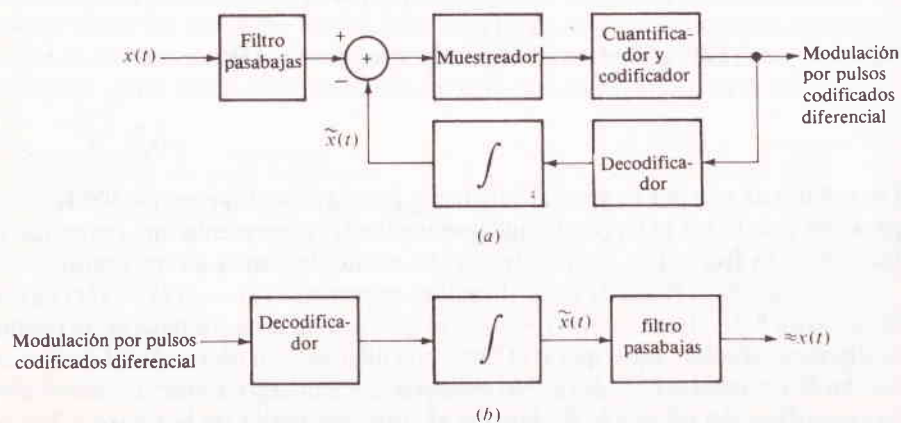



FIGURA 8.22
Modulación por codificación de pulsos diferencial. (*a*) Modulador; (*b*) demodulador.

Modulación por pulsos codificados diferencial ★

Para concluir este tratado sobre los métodos de transmisión digital para señales analógicas, describiremos en forma breve una técnica que combina la estrategia para comparación de realimentación en modulación en delta con la cuantificación multinivel de la modulación por pulsos codificados. Conocida como modulación por pulsos codificados diferencial (DPCM), esta técnica contiene una promesa particular para señales de video.

Funcionalmente, la señal de modulación por pulsos codificados diferencial es una representación de modulación por pulsos codificados de la señal diferencia $x(t) - \tilde{x}(t)$ generada como se muestra en la figura 8.22*a*. Pero $\tilde{x}(t)$ ahora tiene un tamaño variable de escalón que va de $\pm\,\Delta$ a $\pm\,Q\Delta/2$, siendo Q el número de niveles cuánticos, por lo que ésta sigue con más exactitud a $x(t)$. (Si $Q = 2$, la modulación por pulsos codificados diferencial se reduce a modulación en delta.) Así, en especial, si se emplea la compresión-expansión, habrá ruido inactivo mucho más bajo, elevación más rápida, y menos posibilidad de sobrecarga de pendiente. La figura 8.22*b* es el demodulador.

Es obvio que en la modulación por pulsos codificados diferencial con $Q > 2$ se requiere equipo casi tan complejo como en la modulación por pulsos codificados convencional. En cambio, ofrece una reducción potencial del ancho de banda de transmisión. Esto procede puesto que la señal diferencia $x(t) - \tilde{x}(t)$ se representa en forma adecuada con niveles cuánticos mínimos si $x(t)$ no cambia de manera drástica de muestra a muestra. Por ejemplo, la modulación por pulsos codificados diferencial con $Q = 8 = 2^3$ (un código de tres bits) produce una reproducción aceptable de la señal de video, mientras que la modulación por pulsos codificados directa debe tener $Q = 2^8$ para una calidad de imagen monocromática†(blanco y negro) comparable. Así, el ancho de banda de transmisión se reduce en 3/8.

8.4 MULTICANALIZACIÓN POR DIVISIÓN EN TIEMPO

La multicanalización por división en tiempo (TDM) es una técnica para transmitir varios mensajes con facilidad, dividiendo el dominio del tiempo en ranuras, una para cada mensaje.

La esencia de la multicanalización por división en tiempo es muy sencilla, como se ilustra en la figura 8.23. Las diferentes señales de entrada, todas de banda limitada en W por los filtros pasabajas de entrada, están muestreadas en forma secuencial en el transmisor por medio de un interruptor rotatorio o *conmutador*. El interruptor efectúa una revolución completa en $T_s \leq 1/2W$ extrayendo una muestra de cada entrada. Por lo que, la salida del conmutador es una forma de onda de modulación en amplitud de pulsos que contiene las muestras individuales de mensaje entrelazadas en forma periódica en el tiempo. Si hay M entradas, el espaciamiento entre pulsos es $T_s/M = 1/Mf_s$, mientras que el espaciamiento entre muestras sucesi-

† $Q = 2^9$ se emplea para transmisión de TV a color por medio de la modulación por pulsos codificados.

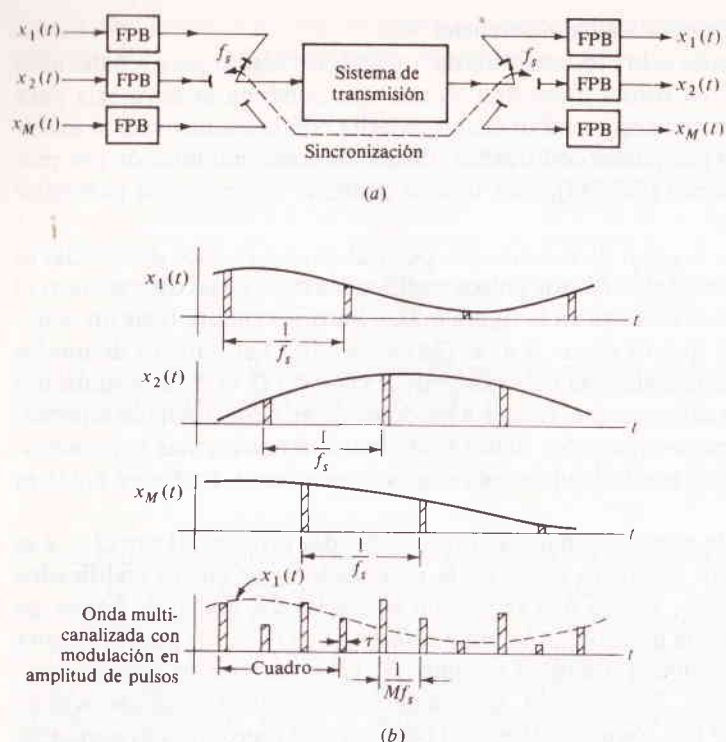



FIGURA 8.23
Sistema de multicanalización por división en tiempo. (*a*) Diagrama a bloques; (*b*) formas de onda.

vas de una entrada cualquiera es, por supuesto, T_s. A un conjunto de pulsos formado de una muestra de cada entrada, se le designa un *cuadro*.

En el receptor, un interruptor rotatorio similar, el *deconmutador* o *distribuidor*, separa las muestras y las distribuye a un banco de filtros pasabajas, el cual a su vez reconstruye los mensajes originales. La acción de conmutación es, por lo general, electrónica, y se proporciona sincronización a las señales para mantener al distribuidor acorde con el conmutador. En efecto, la sincronización es quizá el aspecto más crítico de la multicanalización por división en tiempo.†

Dentro de esta armazón, son muchas las variedades de sistemas de división en tiempo que involucran tanto la modulación pulsada como de CW, y muchas las combinaciones de modulación de subportadora y portadora en los sistemas de división de frecuencia. Los valores muestra entrelazados se pueden convertir en forma directa a modulación en duración de pulsos, por posición de pulsos, o por pulsos codificados y transmitidos con o sin modulación de portadora. En efecto, es la modu-

† Bell Telephone Laboratories (1971, Caps. 25 y 26) cubre lo sorpresivo y notable de la multicanalización digital empleando encuadre estadístico, empaquetado de pulsos y almacenes elásticos. Véase Bennet (1970, Cap. 9) para una comparación más detallada.

lación por pulsos codificados multicanalizada por división en tiempo, la que más promete para emplearse en telefonía. Para propósitos de reducción de ruido, la modulación en posición de pulsos en AM, la modulación por pulsos codificados en FM, etc., resultan particularmente atractivas. Pero antes de entrar en detalles, haremos una pausa para comparar ambos métodos de multicanalización.†

Es obvio que la multicanalización por división en tiempo y por división de frecuencia buscan los mismos objetivos, aunque los medios sean diferentes. Además, a ambas se les puede calificar como técnicas *duales*; porque en la multicanalización por división en tiempo las señales están separadas en el dominio del tiempo, pero mezcladas en frecuencia, mientras que en la multicanalización por división en frecuencia las señales están separadas en el dominio de la frecuencia pero mezcladas en tiempo. Por lo tanto, es razonable preguntarse qué ventajas, si existen algunas, ofrece la multicanalización por división en tiempo en comparación con la multicanalización por división en frecuencia. Desde el punto de vista teórico, ninguna. Desde el punto de vista práctico, la multicanalización por división en tiempo puede ser superior en dos aspectos:

Primero, la instrumentación de este sistema es un poco más sencilla. Recuérdese que la multicanalización por división en frecuencia requiere moduladores de subportadora, filtros pasabanda y demoduladores para *cada* canal de mensaje; todos éstos quedan reemplazados por el conmutador y el distribuidor de la multicanalización por división en tiempo. Y la sincronización en la multicanalización por división en tiempo es ligeramente más exigente que en la multicanalización por división en frecuencia con modulación de portadora suprimida.

Segundo, e igual de importante, la multicanalización por división en tiempo es invulnerable a las fuentes usuales de diafonía intercanal de la multicanalización por división en frecuencia; es decir, al filtraje imperfecto de canal y a la modulación cruzada debidos a las no linealidades. En efecto, no existe diafonía en la multicanalización por división en tiempo si los pulsos están aislados en forma completa y no se traslapan, puesto que la separación de los mensajes la da la deconmutación o selección en tiempo, más que el filtraje. Por lo tanto, la inmunidad a la diafonía de la multicanalización por división en tiempo es contingente con una respuesta de banda ancha y ausencia de distorsión por corrimiento de fase (retardo), manteniendo los pulsos cortos y confinados. (Nótese que la distorsión de fase no origina diafonía en la multicanalización por división en frecuencia.)

Diafonía en la multicanalización por división en tiempo

Las formas reales de los pulsos con colas que decaen, tienden a traslaparse. Sin embargo, la diafonía resultante se puede reducir en forma efectiva dando *tiempos de seguridad* entre pulsos, análogos a las bandas de seguriddad de la multicanalización por división en frecuencia. Así, un sistema práctico de división de tiempo tendrá tantos tiempos como bandas de seguridad, los primeros para suprimir la diafonía, y las últimas para facilitar la reconstrucción de mensajes con filtros prácticos.

† Véase, por ejemplo, Nichols y Rauch (1956).

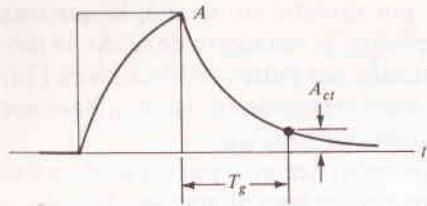



FIGURA 8.24
Diafonía en la multicanalización por división en tiempo.

Para una estimación cuantitativa de la diafonía, se sabe que los pulsos decaen en una forma más o menos exponencial con una constante de tiempo del orden de $1/2\pi B$ en donde B es el ancho de banda a 3 dB del canal, tomado como un canal de banda base. Si T_g es el tiempo de seguridad en el sentido de mínimo espaciamiento entre pulsos, entonces el traslape en el peor de los casos es como se ilustra exagerando, en la figura 8.24. Puesto que $A_{ct} = Ae^{-2\pi BT_g}$, se define el *factor de diafonía*

$$\kappa \triangleq \left(\frac{A_{ct}}{A}\right)^2 = e^{-4\pi BT_g} \qquad (1a)$$

o

$$\kappa \approx -54.5\, BT_g \quad \text{dB} \qquad (b)$$

Así, por ejemplo, si queremos mantener la diafonía abajo de $-$ 30 dB se requiere que $T_g \geq 1/2B$.

Los tiempos de seguridad son importantes especialmente cuando los pulsos multicanalizados son de modulación en duración de pulsos o de modulación en posición de pulsos en vez de modulación en amplitud de pulsos, ya que los pulsos modulados en tiempo "se mueven" en su porción del cuadro.

EJERCICIO 8.7 Se van a transmitir diez señales de voz sobre un canal de banda base de ancho de banda de 400 kHz empleando TDM-PPM. Tomando $\tau = 5\mu s$ y $f_s = 8$ kHz, calcúlese T_g tal que $\kappa = -60$ dB y encuéntrese el máximo desplazamiento permitido t_0 por pulso. *Respuesta*: $T_g \doteq 2.75\ \mu S$, $t_0 = [(1/Mf_s) - \varrho - T_g]/2 = 2.37\ \mu s$.

Filtraje de banda base

Falta un punto más para completar la comparación de la multicanalización por división en frecuencia con la multicanalización por división en tiempo, a saber, el correspondiente a la conservación del ancho de banda. Considérese un sistema de multicanalización por división en frecuencia con M entradas de banda limitada en W. Si las bandas de seguridad son pequeñas en comparación con W y se emplea en todo la modulación de banda lateral única, el ancho de banda de transmisión será MW, lo cual es de manera obvia el mínimo absoluto. Pero, ¿qué hay en relación con la multicanalización por división en tiempo? Podría parecer, por la figura 8.23*b* y por el análisis previo de la modulación en amplitud de pulsos que el ancho de banda de la forma de onda multicanalizada excede mucho al mínimo; es decir, que la duración del pulso es $\tau \ll 1/Mf_s$, por lo que la transmisión como pulsos de RF requiere un ancho de banda de $1/\tau \gg MW$.

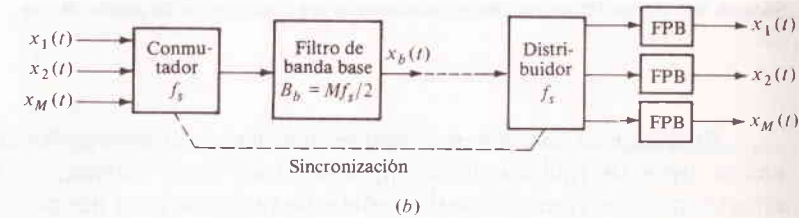
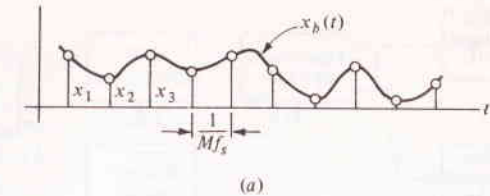



FIGURA 8.25
Filtraje de banda base en la multicanalización por división en tiempo. (*a*) Forma de onda filtrada; (*b*) diagrama del sistema.

Pero nótese que la onda multicanalizada no es otra cosa que una serie de puntos muestra periódicos, aunque de mensajes diferentes. Invertimos el teorema del muestreo, estos puntos pueden ser descritos en forma completa por una forma de onda *continua* $x_b(t)$ que no tiene relación con los mensajes originales salvo que pasa por los valores muestra correctos en los correspondientes tiempos de muestra (Fig. 8.25*a*). Entonces, puesto que los puntos están espaciados en el tiempo en $1/Mf_s$, $x_b(t)$ puede ser de banda limitada en $B_b = Mf_s/2$. En efecto, $x_b(t)$ se obtiene filtrando a pasabajas (banda base) la onda multicanalizada como lo indica la figura 8.25*b*. En el receptor, el distribuidor recoge los puntos de muestra originales de $x_b(t)$ para la reconstrucción en la forma usual.

Si se emplea filtraje de banda base, si la frecuencia de muestreo es cercana a la velocidad de Nyquist ($f_s = 2W$) y si la modulación de portadora es de banda lateral única; el ancho de banda de trasmisión de la multicanalización por división en tiempo resulta ser $Mf_s/2 = MW$. Bajo estas condiciones, la multicanalización por división en tiempo puede lograr el mismo ancho de banda mínimo que la multicanalización por división en frecuencia, pero con pérdida de inmunidad a la diafonía. A los sistemas de multicanalización por división en tiempo de este tipo se les designa modulación en amplitud de pulsos en banda lateral única.

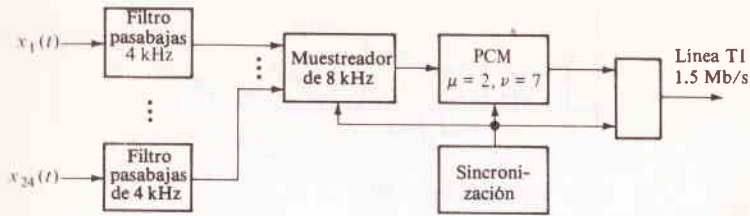



FIGURA 8.26
Sistema Bell de multicanalización en modulación por codificación de pulsos de voz T1.

El filtraje de banda base se emplea más usualmente en aplicaciones que tienen ambos tipos de multicanalización, sobre todo para sistemas de telemetría†. El arreglo común es combinar varias señales de variación lenta por medio de la multicanalización por división en tiempo con filtraje de banda base, formando una onda compuesta, la cual después es multicanalizada por división en frecuencia con otras señales de ancho de banda comparable.

Ejemplo 8.2 Telefonía en multicanalización por división en tiempo con modulación por pulsos codificados

Análogo a su jerarquía de multicanalización por división en frecuencia, el Sistema Bell tiene una jerarquía de multicanalización por división en tiempo *digital* en la cual la rapidez de pulsos binarios o *rapidez de bits* juega el mismo papel que el ancho de banda en la multicanalización por división en frecuencia. Por lo común, los bloques construidos son las "líneas" de T1 a T4 (pares torcidos o cables coaxiales) que tienen rapidez de bits de, en forma aproximada, 1.5, 6.3, 45 y 274 megabits por segundo (Mb/s), respectivamente y sujetas a cambios según los avances tecnológicos. Estas líneas son capaces de transmitir señales digitales de datos o señales analógicas de modulación por codificación de pulsos de diferentes velocidades de bits.

El arreglo de la multicanalización de señal de voz T1 se muestra en el diagrama de la figura 8.26. Se muestrean veinticuatro señales de voz en $f_s = 8$ kHz y la señal resultante de multicanalización por división en tiempo modulación en amplitud de pulsos se convierte a modulación por codificación de pulsos con $\mu = 2$ y $\nu = 7$. Los pulsos binarios adicionales ("bits domésticos") para sincronización y encuadre se insertan elevando la rapidez de salida a 1.544 Mb/s. Se pueden combinar así cuatro señales T1 empleando el multicanalizador M12 para producir una señal T2, y así sucesivamente elevando la jerarquía.

La figura 8.27 es indicadora de la flexibilidad y muestra cómo la voz, datos digitales, teléfono televisivo o videofono y señales de TV a color se pueden combinar para una transmisión sobre una línea T4. Las señales T1 incluyen voz de modulación por codificación de pulsos como se describió antes, y datos digitales multicanaliza-

† Ver por ejemplo, Nichols y Rauch (1956).

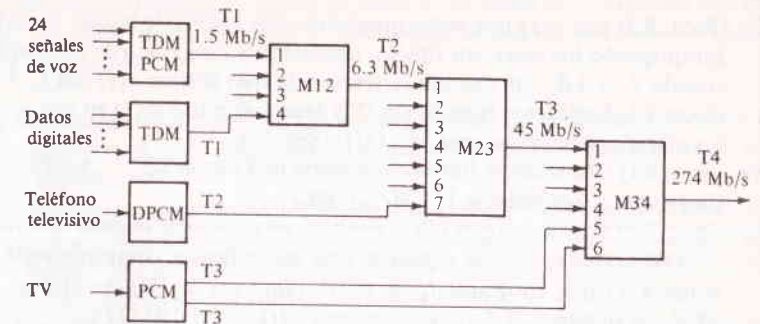



FIGURA 8.27
Configuración ilustrativa de la jerarquía del Sistema Bell de la multicanalización por división en tiempo digital.

dos por división en tiempo. Las señales T2 son T1 multicanalizadas o de teléfono televisivo con modulación por codificación de pulsos diferencial ($f_s \approx 2$ MHz, $\mu = 2$, $\nu = 3$). Un multicanalizador M23 combina siete señales T2 para formar una señal T3. De manera alterna, puesto que la TV a color con modulación por codificación de pulsos binaria requiere 90 Mb/s ($f_s \doteq 10$ MHz, $\nu = 9$), se deben emplear dos líneas T3 para este objeto. Por último, la señal T4 se genera a partir de seis señales T3.

Un quinto bloque construido para la jerarquía está actualmente en desarrollo en la forma de una guía de onda helicoidal operando en la banda de 100 GHz. De resultar factible, esta guía de onda tendrá una capacidad anticipada de 60 × 274 = 16 440 Mb/s, equivalente a, ¡un cuarto de millón de canales bilaterales de voz!

8.5 PROBLEMAS

8.1 (Secc. 8.1) Si se muestrea en forma ideal a $x(t) = \cos 2\pi\, 100t + \cos 2\pi\, 220t$ en $f_s = 300$ y si se pasa a $x_\delta(t)$ a través de un filtro pasabajas ideal con $B = 150$, ¿qué componentes de frecuencia están a la salida? [*Sugerencia*: Bosquéjese el espectro de líneas de lado doble de $x_\delta(t)$.]
Respuesta: 80 y 100 Hz.

8.2 (Secc. 8.1) La señal cuyo espectro se ve en la figura P8.1 se muestra en forma ideal en $f_s = 20$ Hz. Bosquéjese el espectro de $x_\delta(t)$ para $|f| \le 40$ Hz. ¿Se puede recuperar a $x(t)$? De ser posible, ¿cómo? Repítase con $f_s = 30$ Hz.

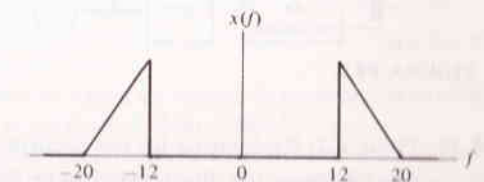



FIGURA P8.1

8.3 (Secc. 8.1) Sea $x(t)$ una señal *pasabanda* con ancho de banda B centrado en $f_0 \gg B$. Bosquejando los espectros típicos, demuéstrese que se puede recuperar a $x(t)$ de $x_\delta(t)$ cuando $f_s > 2B$, aun cuando la frecuencia más alta en $x(t)$ sea $f_0 + B/2 > f_s/2$.

8.4 (Secc. 8.1) Dedúzcase la ecuación (12) desarrollando a $s_\delta(t)$ en una serie exponencial de Fourier y aplicando la ecuación (21), (Sec. 2.4).

8.5▼ (Secc. 8.1) Dedúzcase la fórmula sumatoria de Poisson $\Sigma_{k=-\infty}^{\infty} x(kT_s) = f_s\Sigma_{n=-\infty}^{\infty} X(nf_s)$. (*Sugerencia*: Considérese $\int_{-\infty}^{\infty} x(t)s_\delta(t)dt$.7

8.6★ (Secc. 8.1) Supóngase que $v(t)$ es *limitada en tiempo* de manera que $v(t) = 0$ para $|t| \geq T$. Demuéstrese que su *espectro* está especificado completamente por los valores muestra $V(nf_0)$, en donde $f_0 \leq 1/2T$. {*Sugerencia*: Hágase $z(t) = [v(t)] * [\Sigma\delta(t - kT_0)]$, en donde T_0 3 $1/f_0$, y nótese que $v(t) = z(t)\Pi(t/2T)$.}

8.7 (Secc. 8.1) La señal $x(t) = \text{senc}^2$ está muestreada en forma ideal en $t = 0, \pm 1.0, \pm 0.2, \ldots$, y reconstruida por medio de un filtro ideal con $B = 5$, ganancia unitaria y cero retardo en tiempo. Desarróllese el proceso de recontrucción en forma gráfica, como en la figura 8.5, para $|t| \leq 0.2$.

8.8 (Secc. 8.1) Se muestrea un pulso rectangular con $\tau = 2$ y se reconstruye empleando un filtro pasabajas ideal con $B = f_s/2$. Bosquéjense las formas de onda de salida resultantes cuando $T_s = 0.8$ y 0.4. Supóngase que un tiempo de muestra está en el centro del pulso.

8.9★ (Secc. 8.1) La figura P2.3 es una *retención de primer orden* y tiene $H(f) = T(1 + j2\pi fT)\,\text{senc}^2 f\,T_s^{-j2\pi fT}$. Tomando $T = T_s$, analícese su comportamiento empleando bosquejos como los de la figura 8.7. (*Sugerencia*: Encuéntrese $h(t)$ del diagrama a bloques, no de $H(f)$.)

8.10 (Secc. 8.1) Bosquéjese la función de conmutación para el troceador bipolar de la figura 8.9 y demuéstrese que $s(t) = (4/\pi)\cos\omega_s t - (4/3\pi)\cos 3\omega_s t + (4/5\pi)\cos 5\omega_s t + \ldots$

8.11 (Secc. 8.1) El intervalo útil de frecuencia de un amplificador determinado es f_{co} a $f_{co} + B$. Idéese un método empleando troceadores bipolares para que el amplificador se pueda emplear para una señal que tenga un contenido significativo de CD y $W \gg B$, suponiendo que $B \gg f_{co}$. (*Sugerencia*: Véase el problema 8.10.)

8.12 (Secc. 8.1) Demuéstrese que el sistema troceador de la figura P8.2 combina las operaciones de matrizado y modulación de subportadora y con ello produce la señal de banda base que se requiere para FM estéreo (Fig. 5.37). (*Sugerencia*: Hay, en efecto, dos funciones de conmutación que tienen $\tau = T_s/2$ y difieren en un retardo de $T_s/2$.)

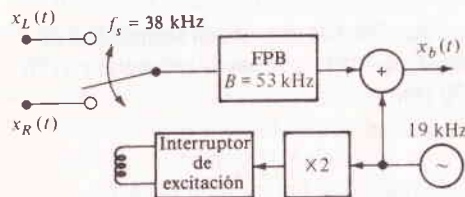



FIGURA P8.2

8.13 (Secc. 8.2) Explíquense los dos siguientes enunciados:

(*a*) La conversión directa a pulsos de RF, en vez de a AM, FM o PM, es la forma más apropiada de modulación de portadora en sistemas de modulación analógica de pulsos.

(*b*) Un sistema de modulación de pulsos en posición de canal único requiere una señal de sincronización, mientras que en la modulación en amplitud de pulsos y en duración de pulsos, no.

8.14 (Secc. 8.2) Encuéntrese $H_{eq}(f)$ para un sistema de modulación en amplitud de pulsos con $p(t) = \cos \pi t/\tau$, $|t| \leq \tau/2$. Bosquéjese la relación de amplitudes sobre $0 \leq f \leq W$ cuando $\tau = 1/5W$.

8.15 (Secc. 8.2) En algunos sistemas de modulación en amplitud de pulsos, el dispositivo de muestreo extrae un valor *promedio*, así, $x(kT_s)$ se sustituye por $\bar{x}(kT_s) \triangleq (1/\tau)\int_{kT_s-\tau}^{kT_s} x(\lambda)d\lambda$. Encuéntrese un filtro de igualación apropiado cuando $p(t) = \Pi(t/\tau)$. [*Sugerencia*: ¿Qué filtro produce $\bar{x}(t)$ cuando $x(t)$ es la entrada?]

8.16 (Secc. 8.2) Una señal de modulación en duración de pulsos tiene $\tau \ll T_s$, y bordes anteriores fijos en $t = kT_s$. Bosquejando formas de onda típicas, demuéstrese que el dispositivo de la figura P8.3 reconstruirá en forma aproximada a $x(t)$ si $RC \gg T_s$ y el interruptor se cierra en forma momentánea cada $t = kT_s$.

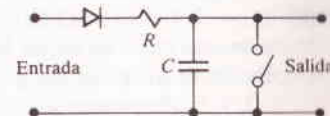



FIGURA P8.3

8.17 (Secc. 8.2) Dedúzcase la ecuación (8) haciendo primero caso omiso de la modulación y escribiendo a $x_p(t)$ como una serie trigonométrica de Fourier similar a la ecuación (2), (Sec. 8.1). Después sustitúyase a τ por $\tau_0[1 + mx(t)]$ y demuéstrese que los términos trasladados en frecuencia tienen la forma

$$(2A_0/n\pi)\,\text{sen}\,\{n\pi\tau_0 f_s[1 + mx(t)]\}\cos n\omega_s t.$$

8.18★ (Secc. 8.2) En la figura P8.4 se muestra el diagrama del *sistema serrasoide de Day* para modulación de fase de banda angosta. La señal de modulación en posición de pulsos se desarrolla empleando un generador de diente de sierra (Figs. 8.12 y 8.13), lo cual explica el nombre del sistema, ya que *serra* es la forma latina de "sierra". Empleando la ecuación (13) demuéstrese que la salida es de la forma $A_c \cos[\omega_c t + \phi_\Delta x(t)]$ cuando el filtro pasabanda está centrado en $f_c = Mf_s$.

$f_c = Mf_s$.

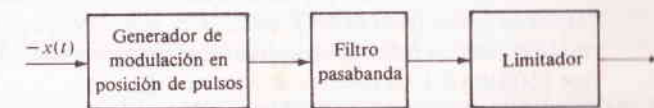



FIGURA P8.4

8.19 (Secc. 8.2) Calcúlese $(S/N)_D$ en términos de γ para un sistema de modulación en amplitud de pulsos que tiene $f_s = 2.5W$, $\tau = 0.2T_s$, $B_\tau = .3/4\tau$, y $m^2x^2 = ½$.

8.20 (Secc. 8.2) Calcúlese γ-ésima y $(S/N)_D$ en términos de γ para un sistema de modulación en posición de pulsos con $\tau = 0.1\,T_s$, $t_0 = 0.25T_s$, y $B_T/W = 50$. Compárese con $(S/N)_{D_{max}}$. *Respuesta*: $(S/N)_D = 20x^2\gamma$.

8.21 (Secc. 8.3) Una señal que tiene un contenido despreciable superior a 3.6 kHz se va a transmitir empleando modulación por pulsos codificados binaria sobre un canal cuya rapidez máxima de pulsos es 40 000 pulsos por segundo. Dibújese un diagrama a bloques del transmisor, especificando valores para todos los parámetros de diseño. *Respuesta*: $f_s = 8$ kHz, $\nu = 5$, $\mathcal{Q} = 32$.

8.22 (Secc. 8.3) Una señal que tiene un contenido despreciable mayor a 20 kHz se va a transmitir empleando modulación por pulsos codificados sobre un canal que tiene $B_T = 70$ KHz. Se requiere que $\mathcal{Q} \geq 60$. Dibújese un diagrama a bloques del transmisor, especificando valores para todos los parámetros de diseño.

8.23 (Secc. 8.3) Generalizando en el ejercicio 8.5, demuéstrese que $\nu \geq \log_\mu 50/P$ cuando el error máximo de cuantificación no es como para exceder el P% de la variación pico a pico de la señal.

8.24 (Secc. 8.3) Considérese una señal con modulación por pulsos codificados binaria, transmitida como pulsos de RF de conexión y desconexión. Si $f_s = 2.5W$ y $\overline{x^2} = ½$ constrúyase gráfica de $(S/N)_D$ en decibeles contra $B_{T\mathrm{RF}}/W$ para $\mathcal{Q} = 4, 16, 64$ y 256. Explíquese por qué esta gráfica es una línea recta.

8.25 (Secc. 8.3) Una señal de voz con un ancho de banda nominal de 3 kHz se va a transmitir sobre un canal de banda base de 20 kHz tal que $(S/N)_D \geq 18\,000_x{}^2$. Dibújese un sistema de modulación por pulsos codificados para efectuar esto, dando valores para μ, ν, $\mathcal{Q}$, y f_s. *Respuesta*: $\mu = 3$, $\nu = 4$, $\mathcal{Q} = 81$, $f_s = 10$ kHz.

8.26 (Secc. 8.3) Un sistema de modulación por pulsos codificados tiene valores fijos de γ y $\mathcal{B}$. Tomando en cuenta el umbral, obténgase una expresión para el valor máximo de $\mathcal{Q}$.

8.27 (Secc. 8.3) Cualesquier muestras cuantificadas sucesivas M se pueden representar por *un* número que tenga $\mathcal{Q}^M$ valores posibles, un proceso al que se designa *hipercuantificación*. Analícese cómo con la modulación por pulsos codificados con hipercuantificación, se puede obtener la *compresión del ancho de banda*, $B_T/W < 1$.

8.28★ (Secc. 8.3) El ruido de cuantificación en la modulación por pulsos codificados se calculó con la ecuación (5) suponiendo que $\epsilon_q(t)$ tiene una función de densidad de probabilidad uniforme, lo cual implica que se presente este mismo caso para $x(t)$ y que los niveles cúanticos estén igualmente espaciados. De manera más general, si b_k es el nivel cuántico k-ésimo y a_k y a_{k+1} son los límites de redondeo adyacentes, entonces

$$\overline{\epsilon_q{}^2} = \sum_{k=1}^{\mathcal{Q}} \int_{a_k}^{a_{k+1}} (x - b_k)^2 p_x(x)\, dx$$

en donde $a_k \leq b_k \leq a_{k+1}$. Demuéstrese que $\overline{\epsilon_q}^2 \approx ⅓Q^2$ cuando hay $Q \gg 1$ niveles espaciados en forma igual en $2/Q$, así $a_k = b_k - 1/Q$ y $a_{k+1} = b_k + 1/Q$ y $p_x(x)$ alisada de modo razonable pero no necesariamente uniforme sobre $[-1, 1]$. (*Sugerencia*: Nótese que $(2/Q)p_x(b_k) \approx$ Prob. $[a_k \leq x \leq a_{k+1}]$.)

8.29★ (Secc. 8.3) con relación al problema 8.28, supóngase que $Q = 2$ y que $p_x(x) = (3/2)(1 - |x|)^2$, $|x| \leq 1$, la cual es una señal con un factor de cresta grande.

(*a*) Calcúlese $\overline{\epsilon_q}^2$ para niveles espaciados en forma igual, es decir, $b_k = -½$ y $½$, $a_k = -1, 0$ y 1.

(*b*) Tomando a $b_k = \pm b$ y a_k como en lo anterior, encuéntrese el valor de b que reduzca al mínimo a $\overline{\epsilon_q}^2$ y compárese el valor resultante con $⅓\, \mathcal{Q}^{\,2} = 1/12$.

8.30★ (Secc. 8.3) Una señal con $W/f_0 = 4$ va a ser transmitida empleando modulación por pulsos codificados binaria o modulación en delta. ¿Sobre qué intervalo de $\mathcal{B}$ dará la modulación en delta el valor mayor de $(S/N)_D$, suponiendo que ambos sistemas están por encima del umbral?

8.31 (Secc. 8.4) Se van a multicanalizar por división en tiempo, veinticinco señales de voz que tienen $W = 3$ kHz, y transmitir empleando modulación en amplitud de pulsos en AM, es decir, modulación de portadora en AM. Dando una banda de seguridad para la reconstrucción de 2 kHz, determínese el mínimo ancho de banda de transmisión requerido y dibújense los diagramas a bloques del transmisor y del receptor.
Respuesta: $f_s = 8$ kHz, $B_T = 200$ kHz.

8.32 (Secc. 8.4) Se van a multicanalizar por división en tiempo cuatro señales limitadas en banda en W, W, $2W$ y $4W$, respectivamente. Idéese una configuración de conmutador tal que cada señal sea muestreada en forma periódica a su propia velocidad de Nyquist y los valores muestra se entrelacen en forma apropiada.

8.33 (Secc. 8.4) Se van a transmitir diez señales, cada una con $W = 4$ kHz, empleando multicanalización por división en tiempo en modulación en posición de pulsos sin modulación de portadora. Tomando $\tau = 1$ y permitiendo un tiempo de seguridad de $2\mu s$ y bandas de seguridad para la reconstrucción de 2 kHz, calcúlese el máximo desplazamiento posible por pulsos. Estímese la reducción resultante en la relación señal a ruido de salida por canal, en comparación con un sistema de modulación en posición de pulsos de canal único que tenga la misma duración de pulso, la misma banda de seguridad, un décimo de la potencia promedio y un décimo del ancho de banda de transmisión.

8.34★ (Secc. 8.4) Se van a transmitir cinco señales analógicas de datos empleando multicanalización por división en tiempo con modulación por codificación de pulsos de banda base con $\mu = 2$ y $\mathcal{Q} = 256$. Las señales no son de banda estrictamente limitada, pero la experimentación ha demostrado que pasando una de ellas a través de un filtro pasabajas Butterworth de tercer orden (Fig. 2.37) con $B = 800$ Hz no ocurre una distorsión significativa. El canal tiene $\mathcal{L} = 60$ dB y $\eta = 10^{-10}$ W/Hz. Suponiendo que la señal binaria entrelazada tiene conformación de pulsos senc como se analizó (Sec. 4.5), dibújese y rotúlese en forma completa el diagrama a bloques del sistema, incluyendo el valor mínimo de S_T.

8.35★ (Secc. 8.4) La diafonía ilustrada en la figura 8.24 es despreciable más allá de la ranura adyacente de tiempo para valores razonables de k. Pero también hay diafonía debida a una transmisión imperfecta de *baja frecuencia*, y ésta puede afectar a varias ranuras de tiempo. Demuéstrese este efecto considerando la respuesta a pulsos rectangulares cuando $H(f) = 1 - \Pi(f/2f_{co})$ en donde $f_0 \ll 1/\tau$. (*Sugerencia*: Aproxímese la forma de onda de salida en vez de hacer el intento de encontrar una expresión exacta.)

8.36 (Secc. 8.4) Analícense las virtudes relativas de la multicanalización por división en tiempo contra la multicanalización por división en frecuencia cuando el canal está sujeto al *desvanecimiento selectivo*, es decir, que una banda angosta de frecuencias experimenta una severa atenuación en comparación con el resto de la banda. Considérense dos casos extremos:

(*a*) Cuando el desvanecimiento es lento y prolongado en comparación con la duración del mensaje.

(*b*) Cuando el desvanecimiento es de corta duración en comparación con el mensaje.

9

TEORÍA DE LA INFORMACIÓN Y LOS SISTEMAS DE COMUNICACIÓN

Los pasados capítulos han tratado de la comunicación eléctrica principalmente en términos de señales, tanto deseadas como indeseables. Se han ideado modelos, examinado los efectos de las redes sobre las señales, y analizado la modulación como un medio para la transmisión de la señal. Aunque se han logrado muchos hallazgos y se ha obtenido mucho conocimiento por este camino, la teoría de la señal sola no es suficiente para una comprensión completa de la materia, en particular cuando se llega al diseño de sistemas nuevos y mejorados. Lo que se necesita es un enfoque que abarque más ampliamente el proceso de la comunicación, una perspectiva más extensa que conduzca a los principios básicos para el diseño y la comparación de los sistemas. En pocas palabras, una teoría general de la comunicación.

Antes de los años cuarentas, se dieron unos cuantos pasos hacia tal teoría, con las investigaciones en el terreno de la telegrafía por parte de Nyquist y Hartley. Pero después, a poco de haber terminado la Segunda Guerra Mundial, Claude Shannon (1948) y Norbert Wiener (1949) expusieron nuevos conceptos que han tenido y continúan teniendo un gran impacto. Tomadas juntas, las ideas de Wiener y Shannon establecieron los cimientos de la moderna *teoría de la comunicación* (estadística). Ambos investigadores se ocuparon de la extracción de la información de un fondo de ruido, y ambos aplicaron conceptos estadísticos al problema.† Hubo, sin embargo, diferencias en los esfuerzos realizados.

† Ambos hombres fueron famosos también por otros logros. Wiener inició la materia conocida como cibernética, y Shannon fue el primero que señaló la relación entre el álgebra y el diseño de circuitos de conmutación, ¡en su tesis de maestría!

Wiener trató el caso cuando las señales que conducen información están más allá del control del diseñador, total o parcialmente y todo el procesamiento se desarrolla en el extremo receptor. Entonces es posible establecer el problema de esta manera: Dado el conjunto de las señales posibles, sin que se puedan elegir, más el inevitable ruido, ¿cómo hacer la mejor estimación de los valores presentes y futuros de la señal que se está recibiendo? Las soluciones óptimas para esto y para problemas similares, se ven en la disciplina conocida como *teoría de la detección*.

El trabajo de Shannon tiene una analogía, más íntima de que se pueda pensar, con la comunicación, donde el procesamiento de la señal puede efectuarse tanto en el transmisor como en el receptor. Shannon estableció este problema: dado el conjunto de mensajes posibles que una fuente puede producir, no de nuestra elección, ¿cómo se representarán los mensajes para que la información sea conducida de la mejor manera sobre un sistema dado, con sus limitaciones físicas inherentes? Para manejar este problema en términos muy generales, es necesario concentrarse más en la información en sí que en las señales, y el enfoque de Shannon fue propio rebautizado como *teoría de la información*.

La teoría de la información es un tema matemático que trata con tres conceptos básicos: la medida de la información, la capacidad de un canal de comunicación para transferir información y la codificación como un medio de utilizar los canales a toda su capacidad. Estos conceptos están ligados en el que se puede designar, teorema fundamental de la teoría de la información, el cual es resumido así:

> Dada una fuente de información y un canal de comunicación, existe una técnica de codificación tal que la información se puede transmitir sobre un canal a cualquier rapidez menor que la capacidad del canal y una frecuencia de errores arbitrariamente pequeña, no obstante la presencia del ruido.

El aspecto sorprendente, casi asombroso, de este teorema es la transmisión *libre de errores* sobre un canal *ruidoso*, una condición que se obtiene por medio del uso de la codificación. En esencia, la codificación se emplea para acoplar la fuente y el canal, a fin de que sea completamente segura la transferencia de información, habiendo en esto una analogía burda en relación con el acoplamiento de impedancia para máxima transferencia de energía.

Pero el estudio de la codificación está, de cualquier manera, fuera del objetivo inmediato en este análisis. Así, con algún desgano, se limita este capítulo en forma primordial a los conceptos de la medida de la información y capacidad de canal, sobre todo, a este último. Haciéndolo así, se llegará de manera eventual a las respuestas a estas ecuaciones significativas:

1. En forma precisa, ¿cómo restringen a la transmisión de información, las limitaciones físicas fundamentales (es decir, el ancho de banda y el ruido?)
2. ¿Existe algo así como un sistema de comunicación *ideal*, y de haberlo, cuáles son sus características?
3. ¿Hasta qué punto los sistemas de comunicación existentes alcanzan el ideal, y cómo pueden mejorar su comportamiento?

Las respuestas a estas preguntas están íntimamente ligadas a la comunicación eléctrica. Serán exploradas con algún detalle al final del capítulo. Pero se debe; empezar con la teoría de la información.

9.1 LA MEDIDA DE LA INFORMACIÓN: ENTROPÍA

El aspecto crucial de la teoría de la información es la medida de la información. Aquí se está empleando la *información* como un término técnico; no se le confunda con sus interpretaciones más convencionales. En particular, la información de la teoría de la información tiene poco qué hacer con el conocimiento o significado, conceptos que constituyen un desafío para una definición precisa, por no mencionar siquiera lo referente a una medida cuantitativa. En el contexto de la comunicación, la información es, en forma simple, aquella que se produce en la fuente para ser transferida al usuario. Esto implica que antes de la transmisión, la información no estaba disponible en el destino; de otra manera la transferencia sería nula. Siguiendo esta línea de razonamiento, considérese la siguiente situación un poco artificiosa.

Un hombre está planeando un viaje a Chicago. Para determinar cuál ropa debe empacar, telefonea a la oficina meteorológica de esa localidad y recibe uno de los siguientes pronósticos:

Saldrá el Sol.
Lloverá.
Habrá un huracán.

Es obvio que la cantidad de información de estos mensajes es muy diferente. El primero, en forma virtual, no contiene información, pues adelanta una seguridad razonable de que el Sol saldrá; no existe incertidumbre acerca de esto, y se ha desperdiciado la llamada. Pero el pronóstico de lluvia proporciona una información que el viajero no poseía antes, porque no llueve todos los días. El tercer pronóstico contiene aún más información, ya que los huracanes son, relativamente, raros e inesperados.

Nótese que los mensajes se han listado en orden decreciente en cuanto a probabilidad y en orden creciente en cuanto a información. El mensaje menos probable, lleva mayor información al usuario. Esto conduce a expresar que la medida de la información está relacionada con la *incertidumbre*, la incertidumbre de parte del usuario acerca de cuál será el mensaje. Además, la cantidad de información depende sólo de la incertidumbre del mensaje, más que de su contenido real o posibles interpretaciones. Si el pronóstico del tiempo en Chicago hubiera sido ''El Sol lloverá huracanes'', conduciría a información, bastante improbable, pero no muy entendible.

De manera alterna, yendo al extremo transmisor de un sistema de comunicación, la medida de la información es una indicación de la *libertad de elección* ejercida por la fuente en la selección de un mensaje. Si la fuente puede escoger libremente entre muchos mensajes diferentes, el usuario tendrá una grande incertidumbre acerca de cuál será el mensaje seleccionado. Pero si no hay posibilidad de elección, si sólo

hay un mensaje posible, no hay incertidumbre y, en consecuencia, tampoco información.

Sea que se prefiera el punto de vista de la incertidumbre o la interpretación de la libertad de elección, es evidente que la medida de la información incluye a las *probabilidades*. Los mensajes de gran probabilidad de elección de parte de la fuente, con conducción de poca información, y viceversa. Esta noción se formaliza definiendo a la autoinformación en términos de probabilidades.

Autoinformación

Considérese una fuente que produce varios mensajes. Desígnese A uno de ellos, y sea P_A la probabilidad de que A se seleccione para su transmisión. Relacionado con el análisis anterior, se escribe la autoinformación asociada con A así:

$$\mathcal{I}_A = f(P_A)$$

donde se va a determinar la función $f(\ )$. Como un paso hacia la obtención de $f(\ )$, el razonamiento intuitivo sugiere que se impongan los siguientes requisitos:

$$f(P_A) \geq 0 \text{ en donde } 0 \leq P_A \leq 1 \tag{1}$$

$$\lim_{P_A \to 1} f(P_A) = 0 \tag{2}$$

$$f(P_A) > f(P_B) \quad \text{para } P_A < P_B \tag{3}$$

Es probable que el estudiante tenga algunas dificultades para interpretar estos requisitos.

Hay muchas funciones que satisfacen la ecuación (1), (2) y 3). El factor final y decisivo proviene de considerar la transmisión de mensajes *independientes*. Cuando se transmite el mensaje A, el usuario recibe $\mathcal{I}_A$ unidades de información. Si se transmite también un segundo mensaje, la información total recibida debe ser la suma de las autoinformaciones, $\mathcal{I}_A + \mathcal{I}_B$. Esta regla sumatoria se aprecia con facilidad si se piensa de A y de B como procedentes de fuentes diferentes. Pero supóngase que ambos mensajes proceden de la misma fuente; se puede entonces hablar de mensaje compuesto $C = AB$. Si A y B son estadísticamente independientes, $P_C = P_A P_B$ y $\mathcal{I}_C = f(P_A P_B)$. Pero la información recibida es aún $\mathcal{I}_C = \mathcal{I}_A + \mathcal{I}_B = f(P_A) + f(P_B)$ y, por lo tanto,

$$f(P_A P_B) = f(P_A) + f(P_B) \tag{4}$$

lo cual es el requisito final para $f(\ )$.

Existe una y sólo una función† que satisface las condiciones planteadas en las ecuaciones de la (1) a la (4), a saber, la *función logarítmica* $f(\ ) = -\log_b(\ )$, en

† Véase Ash (1965, Cap. 1) para la demostración.

donde b es la base logarítmica. Así, la autoinformación se define como

$$\mathcal{I}_A \triangleq -\log_b P_A = \log_b \frac{1}{P_A} \tag{5}$$

en donde b no está especificada por el momento. El signo menos en $-\log_b P_A$ quizá provoque confusión a primera vista. Pero, puesto que las probabilidades están limitadas por $0 \leq P_A \leq 1$, lo negativo del logaritmo es positivo, como se desea. La forma alterna $\log_b(1/P_A)$ ayuda a evitar confusiones en esta contabilidad, y se empleará así en todo.

La especificación de la base logarítmica b es equivalente a seleccionar la *unidad* de información. Mientras que los logaritmos comunes o los naturales ($b = 10$ o $b = e$) parecen ser los candidatos obvios, la convención normal de la teoría de la información es tomar $b = 2$. La unidad de información correspondiente recibe la designación de *bit*, una contracción de *binary digit* (dígito binario) sugerida por J. W. Tukey. Así

$$\mathcal{I}_A = \log_2 \frac{1}{P_A} \qquad \text{bits}$$

El razonamiento en que se apoya ésta un tanto extraña convención es: La información es una medida de la elección ejercida por la fuente; la elección más simple posible es aquella entre dos mensajes equiprobables, es decir, una elección binaria no polarizada. La unidad de información está, por lo tanto, normalizada con esta situación de orden más bajo, y 1 bit de información es la cantidad requerida o conducida por la elección entre dos posibilidades igualmente probables, es decir, si $P_A = P_B = \frac{1}{2}$, entonces $\mathcal{I}_A \mathcal{I}_B = \log_2 2 = 1$ bit.

Los *dígitos* binarios entran en escena por la sencilla razón de que dos cosas cualesquiera se pueden representar por medio de los dos dígitos **0** y **1**. Nótese, sin embargo, que un dígito binario puede conducir más o menos 1 bit de información, dependiendo de las probabilidades. Para prevenir malas interpretaciones, a los dígitos binarios, como elementos de mensaje, se les designa en este capítulo como *binits*.

Puesto que las tablas de logaritmos de base 2 son de manera relativa poco comunes, se requiere de la siguiente relación de conversión:

$$\log_2 v = \log_2 10 \log_{10} v \approx 3.32 \log_{10} v \tag{6}$$

Así, si $P_A = {}^1/_{10}$, $\mathcal{I}_A = 3.32 \log_{10} 10 = 3.32$ bits. En lo que resta de este capítulo, todos los logaritmos serán de base 2, a menos que se indique otra cosa.

Ejemplo 9.1 La información en una imagen

Se dice a menudo que una imagen vale más que mil palabras. Con una ligera ampliación, la medida de la información apoya a este viejo proverbio.

Para efectos de análisis se descompone la figura en un número de puntos discretos, o elementos, teniendo cada elemento un nivel de brillantez variando en pasos del negro al blanco. La imagen normal de televisión, por ejemplo, tiene alrededor de $500 \times 600 = 3 \times 10^5$ elementos y ocho niveles que se distinguen con facilidad. Por lo que hay $8 \times 8x \ldots = 8^{3 \times 10^5}$ imágenes posibles, cada una con probabilidad $P = 8^{-(3 \times 10^5)}$ si se seleccionan en forma aleatoria. Por lo tanto

$$\mathcal{I} = \log 8^{3 \times 10^5} = 3 \times 10^5 \log 8 \approx 10^6 \text{ bits}$$

En forma alterna, suponiendo que los niveles son igualmente probables, la información por elemento es $\log 8 = 3$ bit, para un total de $3 \times 10^5 \times \approx 10^6$ bits, como antes.

Pero ¿qué hay acerca de las mil palabras? Supóngase, por consideración al tema, que un vocabulario lo forman 100 000 palabras igualmente probables. La probabilidad de una palabra cualquiera es entonces $P = 10^{-5}$, por lo que la información contenida en 1 000 palabras es

$$\mathcal{I} = 1{,}000 \log 10^5 = 10^3 \times 3.32 \log_{10} 10^5 \approx 2 \times 10^4 \text{ bits}$$

o sea, sustancialmente menor que la información de una imagen.

La validez de la suposición anterior está, por supuesto, sujeta a controversia; el objeto de este ejemplo es el método, no los resultados.

Entropía y velocidad de información

La autoinformación se define en términos de los mensajes individuales o símbolos que una fuente puede producir. No es, sin embargo, una descripción útil de la fuente en relación con la comunicación. Un sistema de comunicación no está diseñado para un mensaje en particular sino más bien para todos los mensajes posibles, es decir, lo que la fuente podría producir para hacer distinción de lo que ésta produce en una ocasión dada. De esta manera, aunque el flujo instantáneo de información de una fuente puede ser errático, se debe describir a la fuente en términos de la *información promedio* producida. Esta información promedio recibe el nombre de *entropía* de la fuente.

Para una fuente discreta cuyos símbolos son *estadísticamente independientes*, la expresión de la entropía se formula con facilidad. Sea m el número de símbolos diferentes, es decir, un alfabeto de tamaño m. Cuando se transmite el símbolo j-ésimo, éste conduce $j = \log(1/P_j)$ bits de información. En un mensaje largo de $N \gg 1$ símbolos, el j-ésimo ocurre alrededor de NP_j veces, y la información total en el mensaje es aproximadamente

$$NP_1\mathcal{I}_1 + NP_2\mathcal{I}_2 + \cdots + NP_m\mathcal{I}_m = \sum_{j=1}^{m} NP_j\mathcal{I}_j \qquad \text{bits}$$

en la cual, cuando se divide entre N, da la información promedio por símbolo. Por

lo tanto, se define a la entropía de una fuente discreta como

$$\mathscr{H} \triangleq \sum_{j=1}^{m} P_j \mathscr{I}_j = \sum_{j=1}^{m} P_j \log \frac{1}{P_j} \quad \text{bits/símbolo} \tag{7}$$

Se debe observar que la ecuación (7) es un promedio en ensamble. Si la fuente es no estacionaria, las probabilidades de los símbolos pueden cambiar con el tiempo y la entropía tendrá poco significado. Se supondrá en lo sucesivo que las fuentes de información son *ergódicas*, de tal manera que los promedios en tiempo y en ensamble son idénticos.

El nombre *entropía* y su símbolo $\mathscr{H}$ se adoptó de una ecuación similar de mecánica estadística. Dada la similitud matemática, se han hecho varios intentos para relacionar la entropía de la comunicación con la entropía de la termodinámica. † Sin embargo, las relaciones que se han intentado, parecen provocar más confusión que claridad, y quizá es más prudente tratar a las dos entropías como cosas diferentes con el mismo nombre. Por esta razón, se ha sugerido la designación alterna *comentropía* para la entropía de la comunicación.

Pero, ¿cuál es el significado de la entropía de la comunicación como se escribió en la ecuación (7)? Sencillamente esto: aunque no se puede decir cuál será el símbolo inmediato que la fuente produzca, en el promedio se espera obtener $\mathscr{H}$ bits de información por símbolo o $N\,\mathscr{H}$ bits en un mensaje de N símbolos, si N es grande.

Para un tamaño fijo de alfabeto (m fijo) la entropía de una fuente discreta depende de las probabilidades de los símbolos pero está limitada por

$$0 \leq \mathscr{H} \leq \log m \tag{8}$$

Estos límites extremos se interpretan con facilidad y se justifican más allá de toda discusión. El límite inferior, $\mathscr{H} = 0$ implica que la fuente no entrega información (en promedio), y por lo mismo no existe incertidumbre acerca del mensaje. Se podría esperar que esto corresponde a una fuente que produce los mismos símbolos en forma continua; es decir, que toda las probabilidades de los símbolos son cero, salvo para el símbolo que tiene $P = 1$. Se demuestra con facilidad que $\mathscr{H} = 0$ en este caso. En el otro extremo, la entropía máxima debe corresponder a la máxima incertidumbre o a la máxima libertad de elección. Esto implica que todos los símbolos preferidos. Un poco más de reflexión revela que las probabilidades de los símbolos deben ser las mismas, es decir, $\mathscr{H} = \mathscr{H}_{\text{máx}} = \log m$ cuando todas las $P_j = 1/m$, lo cual tiene un significado particular para el trabajo subsecuente.

La variación de $\mathscr{H}$ entre los límites de la ecuación (8) se ilustra mejor considerando una fuente binaria ($m = 2$). Las probabilidades de los símbolos quedan en-

† Véase Brillouin (1956).

tonces relacionadas y se pueden escribir como P y $1 - P$. Así

$$\mathscr{H} = P \log \frac{1}{P} + (1 - P) \log \frac{1}{1 - P} \tag{9}$$

la cual se construye gráficamente conta P en la figura 9.1. Obsérvese el máximo un poco ancho centrado en $P = 0.5$, el caso igualmente probable, donde $\mathscr{H} = \log 2 = 1$ bit.

Si introducimos el elemento tiempo en escena, supóngase que dos fuentes tienen iguales entropías, pero una es "más rápida" que la otra, se producen más símbolos por unidad de tiempo. En un periodo dado, se debe transferir más información de la fuente más rápida que de la lenta, lo cual es obvio que establece mayores demandas en el sistema de comunicación. En esta forma y para estos propósitos, la descripción de una fuente no reside en su entropía sola sino en su *velocidad de entropía*, o velocidad de información, en bits por segundo. La velocidad de entropía de una fuente discreta se define de una manera sencilla como

$$\mathscr{R} \triangleq \frac{\mathscr{H}}{\bar{\tau}} \quad \text{bits/s} \tag{10}$$

donde $\bar{\tau}$ es la duración promedio del símbolo, o sea,

$$\bar{\tau} = \sum_{j=1}^{m} P_j \tau_j \tag{11}$$

por lo que $1/\bar{\tau}$ es igual al número promedio de símbolos por unidad de tiempo.

EJERCICIO 9.1 Calcúlese la velocidad de entropía de una fuente telegráfica que tiene $P_{\text{punto}} = 2/3$, τ_{punto} Δ 0.25, $P_{\text{raya}} = 1/3$, $\tau_{\text{raya}} = 0.4$ s. *Resp.*: $\mathscr{R} = 3.44$ b/s.

Ejemplo 9.2 Entropía y modulación por codificación de pulsos

Como ejemplo de entropía aplicada a lo estudiado apenas, considérese un sistema de modulación por pulsos codificados cuya entrada es la señal continua $x(t)$ de banda

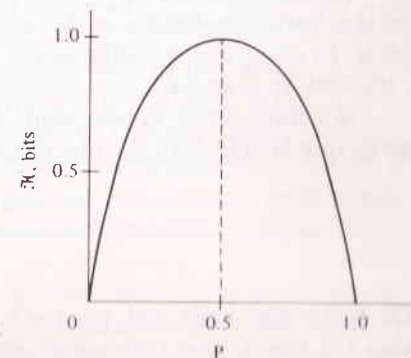



FIGURA 9.1
Entropía de una fuente binaria contra la probabilidad de un símbolo.

limitada en $W = 50$ Hz. Supóngase que se muestra a *x(t)* a la velocidad de Nyquist $f_s = 2W$, y que hay cuatro niveles cuánticos tales que los valores cuantificados tienen probabilidades $^1/_2$, $^1/_4$, $^1/_8$ y $^1/_8$. Identificando cada valor cuantificado posible como un "símbolo", entonces la salida del cuantificador se asemeja a una fuente discreta con $m = 4$ y

$$\begin{aligned}\mathcal{H} &= \tfrac{1}{2}\log 2 + \tfrac{1}{4}\log 4 + \tfrac{1}{8}\log 8 + \tfrac{1}{8}\log 8\\ &= 1.75 \text{ bits/símbolo}\end{aligned}$$

Puesto que la velocidad de los símbolos es $1/\bar{\tau} = f_s = 100$.

$$\mathcal{R} = 100 \times 1.75 = 175 \qquad \text{bits/s}$$

Así, debe ser posible representar esta misma información por medio de dígitos binarios equiprobables (binits) generados a una velocidad de 175 binits/s.

Para comprobar esta idea, supóngase que el sistema es uno de *modulación por codificación de pulsos binaria* con las muestras cuantificadas transmitidas como pulsos binarios codificados. Designando los niveles cuánticos por medio de los números de código 0, 1, 2 y 3, supóngase la siguiente codificación:

Número de código	Probabilidad	Código binario
0	½	**00**
1	¼	**01**
2	⅛	**10**
3	⅛	**11**

Puesto que hay dos dígitos binarios por cada muestra, la señal de modulación por codificación de pulsos tiene 200 pulsos por segundo. Se arguye que deben ser suficientes 175 dígitos binarios por segundo; empero, con este código se requieren 200 por segundo.

La anomalía se resuelve en forma rápida al notar que los dígitos binarios no son igualmente probables; en efecto, $P_0 = {}^{11}/_{16}$ y $P_1 = {}^5/_{16}$, como puede verificar el lector. Es obvio que el código sugerido no es óptimo. Por otra parte, éste es sencillo y eficiente en gran parte.

Si continuamos, la velocidad de los binits y las probabilidades dadas antes sugieren que la velocidad de información a la salida del codificador es

$$\begin{aligned}\mathcal{R} &= 200\left(\tfrac{11}{16}\log \tfrac{16}{11} + \tfrac{5}{16}\log \tfrac{16}{5}\right)\\ &= 200 \times 0.897 = 179 \text{ bits/s}\end{aligned}$$

De nuevo algo anda mal, porque la velocidad de información en el codificador es sólo 175 bits/s. Con toda seguridad que la codificación, ¡no *agrega* información!

Para explicar la discrepancia, se debe recordar que la ecuación de la entropía (7) está basada en los símbolos estadísticamente *independientes*. Y mientras puede ser verdad que las muestras cuantificadas sucesivas son independientes, los pulsos binarios sucesivos no lo son, porque se han codificado en grupos de dos. Para el caso de símbolos independientes, se debe modificar la medida de la información e introducir la entropía condicional.

Entropía condicional y redundancia

Las fuentes discretas a menudo están restringidas por ciertas reglas que limitan las oportunidades en la selección de símbolos sucesivos. La *influencia intersimbólica* resultante reduce la incertidumbre y con ello también la cantidad de información producida. Se percibe este efecto empleando las probabilidades y la entropía condicionales.

La escritura del texto, estando gobernada por reglas de pronunciación y gramática, es un buen ejemplo de la influencia intersimbólica. Sobre la base de la frecuencia relativa, la probabilidad de la U en inglés impresa en $P_U = 0.02$; pero si la letra anterior es Q, la probabilidad condicional de la U, dada la Q, es $P(\text{U Q}) \approx 1$, mientras que, en contraste $P(\text{U W}) < 0.001$. La influencia bien puede extenderse a varios símbolos, frases e incluso oraciones completas, como lo demuestra el hecho de que en muchos textos, éste inclusive, P(QUE SE PUEDE DEMOSTRAR) ≈ 1.

Por lo tanto, la expresión para la entropía condicional se forma considerando la historia transcurrida completa de la fuente (de manera más precisa, todas las posibles historias pasadas). Así, si j representa el símbolo siguiente (o grupo de símbolos) e i representa la secuencia precedente, la información conducida por j dada i es $\log\,[1/P(j|i)]$. Promediando todas las j e i se tiene la entropía condicional

$$\mathcal{H}_c = \sum_i \sum_j P_i\, P(j\,|\,i) \log \frac{1}{P(j\,|\,i)} \tag{12}$$

En general, $\mathcal{H}_c \leq \mathcal{H}$; la igualdad se aplica sólo cuando los símbolos son independientes y $P(j|i) = P_j$.

Se dice de una fuente que produce símbolos dependientes que es *redundante*, significando que esos símbolos son generados, los cuales no son necesarios en forma absoluta para conducir información. (¿Es en realidad necesario escribir de manera explícita la U que sigue a cada Q en inglés?). La redundancia de un texto en inglés se estima que es, en general, del 50%. Esto implica que, al fin y al cabo, la mitad de los símbolos son innecesarios: *yu shld babl t read ths evntho sevrl ltrs r msng* (completo y en español: debe balbucear para leer esto, ya que se han perdido varias letras). El lector quizá desee examinar la observación de que sin redundancia, sería posible emplear abreviaturas, y cualquier arreglo bidimensional de letras formaría un crucigrama válido.

Para pasajes largos de inglés impreso, la entropía condicional puede ser tan baja como de 0.5 a 1.0 bits/símbolo, dadas las inferencias contextuales. Así, con una adecuada codificación, teóricamente, el inglés impreso podría ser transmitido

en forma binaria con un promedio de un dígito binario por símbolo. Esto contrasta con los sistemas de teletipo existentes, los cuales emplean cinco dígitos binarios por carácter.

Desde el punto de vista de una comunicación eficiente, la redundancia en un mensaje es indeseable; se podría enviar la misma información con menos símbolos no redundantes (independientes). Así, la codificación empleada para reducir la influencia intersimbólica, es un método que mejora la eficiencia. Por otra parte la redundancia es un auxiliar definitivo en la resolución de ambigüedades si el mensaje se recibe con *errores*, un fenómeno común en telegrafía, por ejemplo. Además, la codificación para protección de error está basada en la inserción de símbolos redundantes.

Por lo tanto, la transmisión óptima incluye la codificación para reducir la redundancia ineficiente del mensaje, más codificación con el fin de agregar redundancia "eficiente" para control de error. Mucho se ha hecho en el área de la detección de errores y los códigos de detección de errores, pero la codificación para reducir la redundancia en los mensajes es mucho más difícil, y en este sentido se ha hecho relativamente poco. (En forma retrospectiva, el aspecto de reducción del ancho de banda en la modulación por pulsos codificados diferencial se apoya en la redundancia de la señal de entrada.

Fuentes continuas de información

Habiendo examinado las fuentes discretas con cierta extensión, el siguiente paso lógico sería la definición de la entropía para fuentes *continuas*, fuentes cuyos mensajes son funciones del tiempo, variantes en forma continua. Tal definición es posible, pero no se presentará aquí. Por una razón, las matemáticas hacen las cosas más complicadas y tienden a oscurecer la interpretación física; por otra, la entropía de una fuente continua resulta ser una medida relativa en vez de una medida absoluta de la información.

Por fortuna, las metas de este capítulo se pueden lograr adentrándose en la formulación discreta, y muchas de las conclusiones se aplicarán a las fuentes continuas con modificaciones menores. Pero lo más importante, se encontrará que dadas las limitaciones físicas fundamentales, la comunicación es en forma inherente un *proceso discreto* sin importar la fuente. La conclusión contundente es una de las principales contribuciones de Shannon a la teoría de la comunicación, pero ello fue advertido por Hartley mucho antes de 1928.

9.2 CAPACIDAD DE CANAL Y CANALES DISCRETOS

Se ha visto que a menudo es conveniente tratar al equipo terminal de un sistema de comunicación como algo perfecto (libre de ruido, sin distorsión, etc.), y considerar todos los efectos indeseables que ocurren en el canal. Por lo tanto, el canal de comunicación es una abstracción, un modelo que representa al vehículo de transmisión más todos los fenómenos que tienden a restringir la transmisión. El hecho de que

existan limitaciones físicas fundamentales para la transferencia de la información por medios eléctricos, conduce a la noción de *capacidad de canal*.

Tal como la velocidad de la entropía mide la cantidad de información producida por una fuente en un tiempo dado, la capacidad es una medida de la cantidad de información que un canal puede transferir por unidad de tiempo. La capacidad de canal se simboliza por $\mathscr{C}$, y sus unidades son bits/segundo. Reestableciendo el teorema fundamental en términos de $\mathscr{R}$ y $\mathscr{C}$ se tiene:

> Dado un canal de capacidad $\mathscr{C}$ y una fuente con una velocidad de entropía $\mathscr{R}$, entonces si $\mathscr{R} \leq \mathscr{C}$, existe una técnica de codificación tal que la salida de la fuente se puede transmitir sobre el canal con una frecuencia de errores de manera arbitraria pequeña, no obstante, la presencia de ruido. Si $\mathscr{R} > \mathscr{C}$, no es posible transmitir sin errores.

Aunque se intentara hacer plausible el teorema, la demostración completa implica profundizar mucho acerca de la teoría de la codificación y se omite aquí. En vez de ello, se enfocarán aspectos más pertinentes a la comunicación eléctrica.

Capacidad de canal

El teorema fundamental define de manera implícita a la capacidad del canal como la velocidad máxima a la cual el canal surte de información segura al destino. Con esta interpretación en mente se puede formular una expresión general para la capacidad por medio del siguiente argumento.

Considérense todos los mensajes diferentes de longitud T que una fuente podría producir. Si el canal es ruidoso, será difícil decidir en el receptor cuál mensaje en particular fue intentado y se cancela en forma parcial el objetivo de la transferencia de información. Pero supóngase que se restringen los mensajes a sólo aquellos que son "muy diferentes" entre sí, de tal manera que los mensajes recibidos se pueden identificar de manera correcta con una probabilidad de error suficientemente pequeña. Sea $M(T)$ el número de estos mensajes muy diferentes de longitud T.

Ahora, en cuanto se refiere al destino o usuario, a la combinación fuente-canal se le puede considerar como una nueva fuente generadora de mensajes en el extremo receptor. Con la restricción de mensaje anotada antes, esta fuente equivalente es discreta y tiene un alfabeto de tamaño $M(T)$. En correspondencia, la máxima entropía producida por la fuente equivalente es log (MT), y la máxima velocidad de entropía en el destino es $(1/T) \log M(T)$. Por lo que, haciendo que $T \rightarrow \infty$ para asegurar generalidad,

$$\mathscr{C} = \lim_{T \to \infty} \frac{1}{T} \log M(T) \qquad \text{bits/s} \tag{1}$$

una definición alterna de capacidad de canal. El siguiente análisis de canales discretos demuestra que la ecuación (1) es una definición significativamente intuitiva.

Canales discretos sin ruido

Un canal discreto es aquel que transmite información en forma sucesiva, suponiendo diferentes estados eléctricos disjuntos-niveles de voltaje, frecuencia instantánea, etc. Sea μ el número de estados posibles y r la velocidad de señalización en estados por unidad de tiempo. Si la relación señal a ruido es suficientemente grande, la probabilidad de error en el receptor puede ser muy pequeña, tan pequeña que a todos los propósitos se considera al canal como sin ruido. Bajo esta suposición, cualquier secuencia de símbolos se identificará de modo correcto y el cálculo de la capacidad resulta de fiar.

Un mensaje recibido de longitud T consistirá de rT símbolos, siendo cada uno de los cuales uno de los μ estados posibles. El número de mensajes diferentes es así $M(T) = \mu^{rT}$ y en consecuencia

$$\begin{aligned} \mathscr{C} &= \lim_{T\to\infty} \frac{1}{T} \log \mu^{rT} = \lim_{T\to\infty} \frac{rT}{T} \log \mu \\ &= r \log \mu \qquad \text{bits/s} \end{aligned} \tag{2}$$

La capacidad de un canal discreto sin ruido es, por lo tanto, proporcional a la velocidad de señalización y al logaritmo del número de estados. Para un canal binario (μ = 2) la capacidad es numéricamente igual a la velocidad de señalización, o sea, $\mathscr{C} = r$.

De acuerdo con la ecuación (2), se puede duplicar la capacidad del canal duplicando la velocidad de señalización, (lo cual es con certeza razonable) o elevando al cuadrado el número de estados (algo más sutil de apreciar). Atendiendo a lo último, supónganse dos canales idénticos pero independientes operados en paralelo con lo que sus capacidades combinadas son de manera obvia $2(r \log \mu) = 2r \log \mu = r \log \mu^2$. Se puede decir que a la salida se están recibiendo $2r$ símbolos por segundo, extrayéndose cada símbolo de un alfabeto de tamaño μ; o puede verse la salida como r *símbolos compuestos*, extrayéndose a cada uno de un alfabeto de tamaño $\mu \times \mu = \mu^2$.

El siguiente ejemplo expone una técnica de codificación en la que se logra $\mathscr{R} = \mathscr{C}$ en un canal discreto sin ruido.

Ejemplo 9.3

Considérese la fuente de modulación por codificación de pulsos del ejemplo 9.2. En teoría, un canal binario sin ruido bastará para conducir la información si $r \geq 175$ binits/s. Para reducir al mínimo a $\mathscr{R}$ no se puede emplear el código anterior dado que ello requiere de dos dígitos binarios por símbolo de fuente y $r = 200$. Un código más eficiente es el que se muestra a continuación

Número de código	Probabilidad	Código binario
0	½	**0**
1	¼	**10**
2	⅛	**110**
3	⅛	**111**

Con este código, un mensaje que contiene $N \gg 1$ símbolos de fuente, requiere la transmisión de $N/2 + 2(N/4) + 3(N/8) + 3(N/8) = 1.75\,N$ símbolos de canal, o sea, 1.75 dígitos binarios por símbolo de fuente. La capacidad de señalización requerida es entonces $r = 1.75/\tau = 175$, y se ha logrado una transmisión a $\mathscr{R} = \mathscr{C}$; la codificación ha dado lugar a un acoplamiento perfecto entre fuente y canal.

Aunque es de admitirse que este ejemplo es un caso especial, hay dos sugerencias generales en relación con una codificación eficiente que se pueden obtener a partir de éste. Primera, el código es tal, que los símbolos de canal 0 y 1 son igualmente probables y estadísticamente independientes. (El lector escéptico debe verificar esto.) Segunda, al símbolo de fuente con la más alta probabilidad se le asigna el código de palabra más corta, y así, sucesivamente, hasta el símbolo menos probable, al cual se le asigna el código más largo. En la literatura se describen los procedimientos para proyectar tales códigos.†

Entre paréntesis, se podría anotar que el sentido común aconseja asignar las palabras de código más cortas a los símbolos más probables. Hace más de una centuria, mucho antes de Shannon, Samuel Morse construyó su código telegráfico empleando este principio.

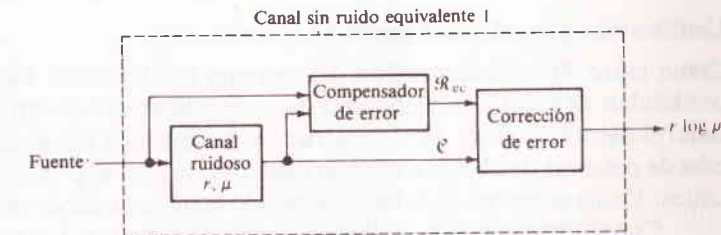



FIGURA 9.2

representando a la letra E con un punto sencillo, etc. A falta de los datos necesarios, Morse estimó las frecuencias de letras tomando en cuenta la distribución de los tipos en el frente de un impresor.

Canales discretos con ruido

Cuando el ruido no se puede menospreciar, la capacidad es menor que $r \log \mu$ a causa de los errores. Se calcula la reducción de la capacidad considerándola como un *compensador de error* ficticio, figura 9.2, el cual examina la entrada y salida del canal y establece las correcciones que se deben hacer. Considérese que el canal está operando a su máximo de $\mathscr{C}$ bits/s, y que la velocidad de información suministrada

† Por ejemplo, Gallager (1968, Cap. 3) o Thomas (1969, Cap. 8).

por el compensador es $\mathscr{R}_{ec}$ bits/s. Entonces, puesto que el canal ruidoso más el compensador son equivalentes al mismo canal sin ruido, la velocidad de información neta sobre el canal ruidoso es $r \log \mu$ menos la velocidad de información del compensador, o sea,

$$\mathscr{C} = r \log \mu - \mathscr{R}_{ec} \qquad (3)$$

Si $\mathscr{R}_{ec} \not\leqslant r \log \mu$, el teorema fundamental afirma que es posible obtener una velocidad diferente de cero, de información en forma virtual exenta de errores, en la salida del canal.

EJERCICIO 9.2 El renombrado *canal simétrico binario* es aquel en el cual ambos estados tienen la misma probabilidad de error, digamos p. Demuéstrese que $\mathscr{R}_{ec} = r\{p \log (1/p) + (1 - p) \log [1/(1 - p)]\}$ y por ello

$$\mathscr{C} = r[1 + p \log p + (1 - p) \log (1 - p)] \qquad (4)$$

Nótese que $\mathscr{C} = 1$ si $p = ½$. ¿Por qué? (*Sugerencia*: El compensador de error produce dos mensajes posibles, "Error" y "Sin error".)

Codificación para el canal simétrico binario

Como parte de su demostración del teorema fundamental, Shannon comprobó la posibilidad de una transmisión muy cercana a la exenta de errores a $\mathscr{R} \leq \mathscr{C}$ en un canal simétrico binario. Su demostración, bosquejada en seguida, incluye la selección de palabras de código en *forma aleatoria*, la cual al principio se enfrentó con la crítica. Desde entonces, se le ha considerado como un método de gran trascendencia.

Considérese que el canal tiene por conveniencia $r = 1$ y que $q = 1 - p$, así, de la ecuación (4),

$$\mathscr{C} = 1 + p \log p + q \log q \qquad (5)$$

Supóngase que la fuente tiene m mensajes equiprobables, cada uno de los cuales se representa por una palabra de código con N dígitos binarios. Esto requiere que $\bar{\tau} = N$ y, por lo tanto,

$$\mathscr{R} = \frac{\log m}{N} \qquad (6)$$

puesto que $\mathscr{H} = \mathscr{H}_{\text{máx}}$.

A causa del ruido, las palabras de código recibidas tendrán errores y, por lo que se sabe acerca de la distribución binomial, el número de errores esperado por palabra es $\bar{n} = Np$. Pero esto no significa de manera necesaria que el mensaje decodificado será erróneo porque, sin $m \ll 2^N$, se pueden seleccionar las palabras de código m de tal manera que ellas sean "muy diferentes" unas de otras y reconocidas en forma correcta no obstante los errores. A modo de ilustración, supóngase que $m = 2$ y $N = 3$; hay $2^3 = 8$ palabras de código binarias de tres dígitos posibles, pero sólo se necesitan $m = 2$ de éstas. Si las palabras seleccionadas son 000 y 111, un error único se corrige con facilidad en forma automática empleando la regla de la pluralidad, por ejemplo, se decodifica a 001 como 000 y a 101 como 111. Generalizando, si se seleccionan las palabras de código en tal forma que difieran entre sí por lo menos $\bar{n} + 1$ dígitos; entonces, la probabilidad de errores por decodificación resulta ser, por pequeña, despreciable siempre y cuando $N \gg m$.

Sin embargo, Shannon dijo que las palabras de código debían seleccionarse de manera aleatoria cuando $N \gg 1$ y $p < ½$. (Si $p > ½$, ¡sólo se intercambian los 1 y los 0 en el receptor!). Bajo estas condiciones, demostró que la probabilidad de un error de decodificación es

$$P_e \approx \frac{m}{2^N} p^{-pN} q^{-qN} \sqrt{\frac{pN}{2\pi q}} \qquad (7)$$

lo cual aproxima la probabilidad de que dos o más palabras difieran en Np dígitos o menos. Entonces, multiplicando la ecuación (5) por $-N$ se tiene $2^{-N\mathscr{C}} = 2^{-N}p^{-pN}q^{-qN}$ (una manipulación muy hábil) por lo que

$$P_e \approx m 2^{-N\mathscr{C}} \sqrt{\frac{pN}{2\pi q}} \qquad (8)$$

Por lo tanto, imponiendo la condición

$$m \leq \frac{2^{N\mathscr{C}}}{N^{\alpha}} \qquad \alpha > ½ \qquad (9)$$

y haciendo que $N \to \infty$, se tiene

$$P_e \leq \lim_{N \to \infty} \frac{1}{N^{\alpha - ½}} \sqrt{\frac{p}{2\pi q}} = 0 \qquad (10)$$

Por último, insertando la ecuación (9) en la (6),

$$\mathscr{R} \leq \lim_{N \to \infty} \frac{1}{N} \left[N\mathscr{C} - \left(\frac{\alpha}{N}\right) \log N \right] = \mathscr{C} \qquad (11)$$

lo cual da como resultado una transmisión exenta de errores en $\mathscr{R} \leq \mathscr{C}$, por lo menos en el límite cuando $N \to \infty$.

Hablando en forma más práctica, parece razonable inferir que se puede acercar de manera arbitraria a esta situación por medio de una cuidadosa elección de las palabras de código y con N grande pero finita. Las técnicas para la codificación de control de error prácticas se examinan en el capítulo 10. Por ahora, se enfoca la atención al caso de los canales continuos.

9.3 CANALES CONTINUOS

Un canal continuo es aquel en el cual los mensajes se representan como formas de onda, es decir, funciones continuas del tiempo, y los parámetros apropiados son el ancho de banda B y la relación señal a ruido S/N. Por sencillez, se tratará sólo un canal de banda base cuya respuesta de frecuencia se ha igualado para hacer la plana sobre $|f| \leq B$. Aunque el canal es continuo y tiene ruido, resulta instructivo desarrollar una relación intuitiva para r y μ, los parámetros de un canal discreto sin ruido.

En relación con la velocidad de señalización r, se ha demostrado ya en la sección 4.5 que $r \leq 2B$ en conde la igualdad incluye el empleo de pulsos senc. Como al número equivalente de "estados" de canal (es decir, niveles de voltaje, etc.), no existe limitación inherente para μ en ausencia de ruido; los niveles de voltaje que tienen espaciamiento arbitrariamente pequeño aún se podrían distinguir en la salida. El valor de μ en un canal ruidoso se estima en términos de la relación señal a ruido, de la siguiente manera.

Sean S y N la potencia promedio de la señal y la potencia de ruido en la salida del canal respectivamente, por lo que la potencia total recibida es $S + N$, y el voltaje rms de salida es $\sqrt{S + N}$. Por razón de la contaminación por ruido, nunca se puede identificar con exactitud el voltaje de señal propuesto. Pero se le puede identificar con una razonablemente baja probabilidad de error si los niveles de voltaje están separados en una cantidad igual o que sobrepase al voltaje rms σ de ruido. Así, en el extremo receptor se tienen niveles de voltaje *espaciados en* $\sigma = \sqrt{N}$ y un *intervalo* de voltaje rms de $\sqrt{S + N}$. El número máximo de estados de canal es, por lo tanto, en forma aproximada

$$\mu = \frac{\sqrt{S+N}}{\sqrt{N}} = \left(1 + \frac{S}{N}\right)^{1/2}$$

Por ejemplo, un canal binario requeriría $(1 + S/N)^{1/2} = 2$, o $S/N = 3$; con ruido gaussiano, la probabilidad de error correspondiente es de alrededor de 0.05.

Combinando estos valores de r y μ se tiene

$$\mathscr{C} = r \log \mu = 2B \log \left(1 + \frac{S}{N}\right)^{1/2}$$

Por lo que,

$$\mathscr{C} = B \log \left(1 + \frac{S}{N}\right) \quad \text{bits/s} \qquad (1)$$

Esta famosa ecuación es conocida como la *ley de Hartley-Shannon*. (Hartley hizo el trabajo preliminar y Shannon la dedujo rigurosamente.) Está escrita en términos de parámetros que se aplican lo mismo a canales discretos que a continuos, sugiriendo

que la capacidad de un canal continuo es $B \log (1 + S/N)$. En efecto, la deducción de Shannon se basó en el caso continuo y se examinará en breve.

La ley de Hartley-Shannon, con el teorema fundamental, tiene dos implicaciones importantes para los ingenieros en comunicaciones. Primera, expresa la forma óptima absoluta con que se puede obtener una transmisión de información segura, dados los parámetros del canal. Segunda, en caso de una velocidad de información específica, expresa que se puede reducir la potencia de la señal, siempre y cuando se incremente el ancho de banda en una magnitud apropiada, y viceversa.

El intercambio de ancho de banda por potencia o relación señal a ruido no es nuevo, porque se ha notado el efecto en sistemas con reducción de ruido de banda ancha como la FM y la modulación por pulsos codificados. Pero la ley de Hartley-Shannon especifica el intercambio posible *óptimo* y además implica que *es posible la comprensión* del ancho de banda. Para aclarar, supóngase que se desea transmitir datos digitales a una velocidad de 30 000 bits/s. De acuerdo con la teoría, se podría emplear un canal que tenga $B = 30$ kHz y $S/N = 1$, puesto que

$$\mathscr{C} = 30 \times 10^3 \log (1 + 1) = 3 \times 10^4 \text{ bits/s}$$

En forma alterna, se puede reducir el ancho de banda a $B = 3$ kHz si la potencia se incrementa en un factor de 1 000, o sea, $S/N = 10^3 = 30$ dB. (Nótese la aproximación $10^3 \approx 2^{10}$.) En forma incidental, los últimos parámetros son típicos de circuitos telefónicos de voz normales; pero cuando se emplean para señales digitales, la velocidad de los datos es normalmente de 4 800 bits/s o menor, indicando un considerable espacio para un empleo ventajoso.

Algunos de los temas anteriores serán examinados más a fondo en la sección 9.4 después de analizar el canal continuo en forma más minuciosa. Puesto que prácticamente todos los sistemas de comunicación son capaces de manejar señales continuas y todos los sistemas tienen ruido, el canal continuo ruidoso amerita una investigación detallada. Como preliminar, se describirán las características de un sistema de comunicación ideal con un canal continuo.

EJERCICIO 9.3 Si $(S/N) \gg 1$, demuéstrese que el tiempo mínimo que se requiere para transmitir K dígitos binarios es

$$T_{\min} \approx \frac{3K}{B(S/N)_{\text{dB}}} \qquad (2)$$

(*Sugerencia*: Véase la ecuación (6), (Sec. 9.1).)

Sistemas de comunicación ideales

Un sistema de comunicación capaz de transmitir sin errores a una velocidad de $B \log (1 + S/N)$ bits/s, en donde B y S/N son los parámetros del canal, recibe la designación de sistema *ideal*. Mientras que un sistema práctico no es o no puede ser ideal, es posible visualizar sistemas cuyo comportamiento se acerque a esa condición. Como

ayuda en el diseño de tales sistemas, se examinan las características de un sistema muy cercano al ideal.

Para empezar, el número de señales diferentes (o mensajes o símbolos) de longitud T es $M = (1 + S/N)^{BT}$, y la máxima velocidad de información es $\mathcal{R} \mp (1/T) \log M$. Las señales del canal se escogen de manera que se puedan identificar en el receptor con una muy pequeña probabilidad de error. Shannon demostró que si las señales en sí son funciones muestra de ruido gaussiano seleccionadas en forma aleatoria y que si $2BT \gg 1$, entonces esta condición resulta bastante aproximada.

La información producida por una fuente es conducida a través del sistema de la siguiente manera: se observa la salida de la fuente por T segundos, y el mensaje se representa (codificado) como una de las señales del canal parecidas a ruido, que se transmite después; así, la información se codifica en *bloques* de longitud T. (Esto implica que el número de mensajes de fuente posibles de longitud T no es mayor que M). En la salida, la señal recibida más ruido se compara con las copias almacenadas de las señales del canal.† La que mejor acopla la señal más ruido se presume que es la señal que en realidad se transmite, y el mensaje correspondiente se decodifica. Se requiere un retardo total en tiempo de $2T$ para las operaciones de codificación y de decodificación.

En la descripción anterior se ha supuesto en forma tácita que $T \to \infty$, porque sólo en este límite se satisfacen todas las condiciones, de tal manera que $\mathcal{R} = B \log(1 + S/N)$. Así, las características de un sistema ideal son como sigue:

1 La velocidad de información se acerca a $B \log(1 + S/N)$.
2 La frecuencia de errores se aproxima a cero.
3 Las propiedades estadísticas de la señal transmitida se aproximan a las del ruido blanco gaussiano de banda limitada.
4 El retardo en tiempo de codificación se incrementa en forma indefinida.

De manera adicional, como se demostrará, el sistema tiene un efecto de umbral agudo.

Por supuesto, que si el ancho de banda del canal es suficientemente grande, se puede aún tener $2BT \gg 1$ con valores razonables de T. Sin embargo, el diseño de los sistemas cercanos al ideal no es un asunto trivial, porque se debe desbalancear el retardo en la codificación y la selección de la señal contra la seguridad. Se arguye una identificación precisa de la señal por una T grande. Pero conforme T aumenta, M debe aumentar en forma exponencial para mantener constante la velocidad de información. Rice Demostró (1950) que para lograr $\mathcal{R} = 0.96\mathcal{C}$ con $S/N = 10$ y una probabilidad de error de 10^{-5}, el número de señales de canal que se requiere ¡es $M = 2^{10\,000}$!

Es obvio que una *codificación de bloque* eficiente con su enorme número de señales de canal es casi tan impráctica como el retardo de codificación infinito de un sistema ideal. Los esquemas alternos que emplean la *codificación secuencial* son casi tan eficientes y requieren de un equipo relativamente sencillo.

† El Apéndice A comprende la implementación de este proceso de comparación bajo el encabezado de Teoría de la Detección.

Descripción del espacio de la señal en la comunicación

No hace mucho que los números imaginarios eran juguetes de las matemáticas puras, juzgados como sin valor práctico. Pero los físicos e ingenieros electricistas desde entonces les han asignado una útil interpretación como $\sqrt{-1}$, y en la actualidad es casi inimaginable examinar el análisis de señales, las ondas electromagnéticas, o la teoría de sistemas sin la ayuda de esta herramienta. De manera similar, la una vez esotérica geometría de los espacios multidimensionales (*hiperespacio*) tuvo un nuevo significado debido a Shannon (1949) para el estudio de la comunicación en canales continuos, reduciendo un problema que de otro modo hubiera sido intratable, a términos más familiares. En esta sección se bosqueja su deducción de la ley de Hartley-Shannon, empleando los conceptos del espacio de la señal.

Considérese un canal continuo de banda base con un ancho de banda B de tal manera que, por necesidad, todas las señales en el extremo receptor son de banda limitada en B. Recurriendo a la teoría del muestreo —a saber, la ecuación (15), (Sec. 8.1)—, cualquiera de las señales de información se puede escribir como

$$x(t) = \sum_k x_k \operatorname{senc}(2Bt - k) \tag{3}$$

en donde

$$x_k \triangleq x(kT_s) \qquad T_s = \frac{1}{2B}$$

Se supone que estas señales se extraen de un ensamble ergódico con una restricción de potencia promedio; así,

$$S = \overline{x^2} = \overline{x_k^{\,2}} \tag{4}$$

como se desprende de la ecuación (12), (Sec. 4.5).

Ahora, en forma temporal, supóngase que *x(t)* es esencialmente cero fuera de un intervalo de tiempo "largo" de duración T; entonces se le describe en forma completa por medio de $D = T/T_s = 2BT$ valores muestra, $x_1, x_2, \ldots, x_D$. (Cierto una señal de banda limitada no puede ser en forma simultánea en tiempo, y eventualmente se hará que $T \to \infty$ para compensar esto.) El hecho de que D números especifiquen de manera única a *x(t)* conduce a la noción de *espacio de la señal*, un espacio D-dimensional en el cual se representa a *x(t)* por medio de un *vecto* (o D-ada)

$$x = (x_1, x_2, \ldots, x_D) \qquad D = 2BT \tag{5}$$

El vector empieza en el origen y termina en un punto cuyas coordenadas son $x_1, x_2, \ldots, x_D$.

El espacio D-dimensional es como el espacio ordinario salvo porque tiene D ejes mutuamente perpendiculares. Y no obstante que no se puedan construir más de tres ejes de tales condiciones en nuestro mundo tridimensional,† se puede tratar en

† Abbott (1950) es una divertida plática acerca de los misterios del espacio tridimensional como lo vio un nativo de Planolandia, un espacio bidimensional.

forma lógica y matemática con espacios de mayor dimensionalidad. En particular, el espacio de la señal es *euclidiano*, es decir, que el cuadrado de la distancia del origen a cualquier punto es la suma de los cuadrados de las coordenadas. La magnitud o norma al cuadrado de un vector de señal es, por tanto

$$\|x\|^2 = x_1^2 + x_2^2 + \cdots + x_D^2 = \sum_{k=1}^{D} x_k^2 \tag{6}$$

Si $T \gg 1/2B$, lo cual cuantifica el significado del intervalo "largo", entonces $D \gg 1$ y

$$\frac{1}{D}\sum_{k=1}^{D} x_k^2 \approx \overline{x_k^2} \tag{7}$$

Por lo que, combinando las ecuaciones (4), (6) y (7)

$$\|x\| = \sqrt{DS} = \sqrt{2BTS} \tag{8}$$

la longitud de todos los vectores de señal es proporcional a la raíz cuadrada de la potencia promedio de la señal.

Si el extremo de x se desplaza alcanzando todas las posiciones posibles, la superficie generada así es una *hiperesfera* de radio $\|x\|$ y todos los vectores de señal posibles terminan en la superficie de esta hiperesfera. El "volumen" encerrado por tal esfera es‡

$$\mathscr{V}_D = K_D \|x\|^D \tag{9}$$

en donde la constante K_D no es en forma particular de la incumbencia de este enfoque. Una consecuencia curiosa de la ecuación (9) es que la mayor parte del volumen de una hiperesfera de alta dimensionalidad ($D \gg 1$) se concentra en la superficie. Como aclaración, el volumen relativo entre $\|x\|/2$ y x es $1 - 2^{-D}$, por lo que si $D = 3$ (una esfera convencional), el 87.5% del volumen está en la "mitad" exterior; si $D = 100$, el volumen relativo de la porción exterior es de aproximadamente $1 - 10^{-30}$. Este efecto de *concentración de volumen* demuestra ser útil en este desarrollo porque la dimensionalidad de los espacios de señal típicos es además grande. Por ejemplo, una llamada telefónica de tres minutos con $B = 4$ kHz tiene $D \approx 10^6$.

Nuestra descripción vectorial de las señales de canal también se aplica al *ruido* siempre y cuando sea ruido blanco gaussiano de una fuente ergódica, de banda limitada en B. Con esta condición, los valores muestra espaciados en $1/2B$ son no correlacionados y estadísticamente independientes. La energía de ruido en el tiempo T es entonces muy cercana a NT, siendo N la potencia promedio de ruido. Por lo tanto, el ruido se representa en el espacio de la señal por medio de un vector de longitud $\sqrt{DN}$, y todas las señales posibles de ruido están contenidas dentro de una esfera de ese radio. Dado que el ruido es *aleatorio*, podría parecer que la esfera de ruido debe estar "velluda"; es decir, una función muestra de longitud T puede tener una

‡ Sommerville (1929).

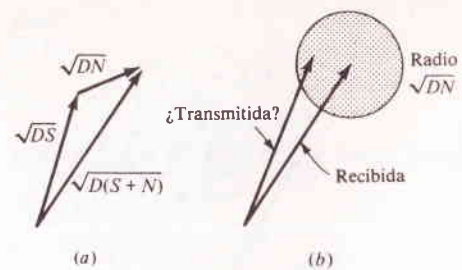



FIGURA 9.3
Representaciones del espacio de la señal. (*a*) Vectores de señal y ruido; (*b*) esfera de incertidumbre debida al ruido.

energía muy diferente de NT. Pero si la dimensionalidad es alta, la concentración de volumen indica que la esfera de ruido está definida en forma muy aguda, más como una pelota de ping pong que como una nube de gas.

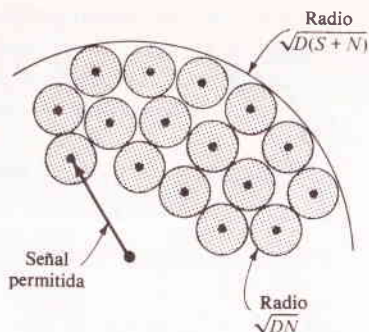
Considérese ahora el estado de cosas en la salida del canal, donde se tiene la señal deseada contaminada con ruido. Bajo la suposición usual de que la señal y el ruido son independientes, sus potencias promedio se suman, y el vector de señal más ruido tiene longitud $\sqrt{D(S + N)}$. La figura 9.3*a* muestra la interpretación geométrica de la señal transmitida, ruido agregado y la señal más el ruido. Se ve que la señal transmitida cae dentro de una esfera de radio $\sqrt{DN}$ en el extremo del vector de la señal más ruido (Fig. 9.3*b*), y esta esfera de ruido indica la incertidumbre del receptor respecto a cuál señal fue destinada.

Si las señales transmitidas posibles se conocen de antemano en el receptor, y si la esfera de incertidumbre contiene el extremo de uno y sólo uno de los vectores de señal posibles, entonces la señal intentada se puede *determinar de manera exacta*, no obstante el ruido. Así, supóngase que se pone dentro de la hiperesfera de radio $\sqrt{D(S + N)}$ un número grande de esferas de ruido que no se traslapan de radio $\sqrt{DN}$ y después se envían sólo aquellas señales correspondientes a los puntos centrales de las esferas de ruido (Fig. 9.4). Cuando se seleccionan de esta manera las señales transmitidas, es posible conducir información sobre un canal continuo ruidoso con una probabilidad de error pequeña que tiende a desaparecer.

¿Cuántas esferas pequeñas de ruido se pueden empacar dentro de la esfera grande de señal más ruido sin que se traslapen? El cálculo es importante, porque expresa que $M = M(T)$, el número de señales (mensajes) "muy diferentes" de longitud T que se pueden identificar de una manera correcta en la salida del canal, del cual se puede encontrar después la capacidad. Es obvio que M no excede el volumen de la esfera grande dividido entre el volumen de una de las esferas pequeñas, es decir, empleando la ecuación (9),

$$M \le \frac{K_D[\sqrt{D(S + N)}]^D}{K_D[\sqrt{DN}]^D} = \left(1 + \frac{S}{N}\right)^{D/2} \tag{10}$$

Obsérvese que M es finito para todos, excepto para los canales verdaderamente exentos de ruido. Además, dado que un canal real tiene ruido y que M es finito, una

FIGURA 9.4
Vectores de señal para transmisión de modo virtual sin errores.

comunicación sobre un canal continuo es en forma inherente un proceso *discreto*. El poner $D = 2BT$ e insertar a M en la ecuación (1), (Sec. 9.2), da

$$\mathscr{C} \leq \lim_{T\to\infty} \frac{1}{T} \log \left(1 + \frac{S}{N}\right)^{BT} = B \log \left(1 + \frac{S}{N}\right) \quad (11)$$

lo cual es un límite superior en la capacidad de un canal continuo.

Para demostrar que la información se puede transmitir con efectividad en $\mathscr{R} \leq B \log(1 + S/N)$ con errores despreciables, Shannon propuso la selección de las M formas de onda *de manera aleatoria*. Si se envía una forma de onda o señal en particular y da por resultado el vector de señal más ruido recibidos cuyo diagrama aparece en la figura 9.5, no habrá confusión ni error de "decodificación" siempre y cuando todos los otros $M - 1$ vectores de señal caigan fuera del volumen en forma de lente indicado. En consecuencia, puesto que las señales se escogen de modo aleatorio, la probabilidad de error P_e es igual a $(M - 1)$ veces la relación del volumen en forma de lente al volumen de la esfera de la señal.

El volumen de una lente D-dimensional resulta difícil de calcular, pero es obviamente menor que el volumen de una esfera de radio h donde, por simple

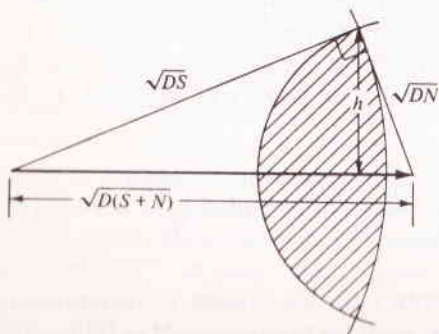



FIGURA 9.5

geometría $h = \sqrt{DSN/(S+N)}$. Por lo tanto,

$$P_e \leq (M-1)\, \frac{K_D h^D}{K_D \|x\|^D} = (M-1) \left(\frac{N}{S+N}\right)^{D/2} \quad (12)$$

por lo que a P_e se le puede hacer tan pequeña como se desee si

$$M - 1 \leq \left(1 + \frac{S}{N}\right)^{BT} P_e \quad (13)$$

similar a la ecuación (9), (Sec. 9.2). Por último, tomando el logaritmo de la ecuación (13), se tiene

$$\frac{1}{T} \log (M-1) \leq B \log \left(1 + \frac{S}{N}\right) - \frac{1}{T} \log \frac{1}{P_e}$$

Por lo que, se puede elegir a M tal que $\mathscr{R} = (1/T) \log M$ se aproxime en forma arbitraria a un valor próximo a $\mathscr{C} = B \log(1 + S/N)$ en el límite conforme $T \to \infty$. Además puesto que la ecuación (12) supone señales seleccionadas en forma aleatoria, debe haber algunos conjuntos específicos de M señales que originen una probabilidad de error aún menor.

Efecto de umbral y modulación de banda ancha

Es importante observar que la obtención de transmisión de información a una velocidad de $B \log(1 + S/N)$ bits/s requiere que $D = 2BT \gg 1$ y $M(T) = (1 + S/N)^{BT}$; en otras palabras las esferas de ruido se deben empacar tan juntas como sea posible sin que se traslapen. Ahora supóngase que la relación señal a ruido cae ligeramente abajo del valor de *diseño*. Las esferas de ruido entonces se traslaparán y el receptor tendrá frecuentes errores de decodificación. Por ello, hay un *efecto de umbral* agudo en el que un pequeño incremento de la potencia de ruido (o una disminución de la potencia de la señal) produce un incremento grande en la probabilidad de error. Como resultado de estos errores, se pierde la información.

Con bastante sorpresa se comprueba que esta misma explicación es válida para el efecto de umbral en la transmisión analógica de mensajes que emplea la modulación banda ancha, por ejemplo, la FM y la modulación por codificación de pulsos. La descripción del espacio de la señal de la modulación es un *mapeo uno a uno* del espacio de los vectores de mensaje dentro del espacio de los vectores de señal modulada; si la modulación es de banda ancha ($B_T \gg W$), el mapeo es entre espacios de diferente dimensionalidad. Pero un teorema muy conocido de topología dice que cualquier mapeo uno a uno entre espacios de diferente dimensionalidad debe ser *discontinuo* en el sentido de que una trayectoria continua en un espacio se mapea en una trayectoria interrumpida en el otro. Por lo tanto, los vectores adyacentes en el espacio de la señal modulada no representan necesariamente vectores adyacentes en el espacio del mensaje, y un ligero traslape de las esferas de ruido puede ocasionar que el receptor "demodule" un mensaje totalmente diferente del que fue enviado.

De conformidad, se concluye que la *mutilación y el efecto de umbral son inevitables en todos los tipos de modulación de banda ancha*.

9.4 COMPARACIÓN DE SISTEMAS

La ley de Hartley-Shannon se aplica a un reducido tipo de canales, a saber, los canales continuos que tienen una restricción de potencia promedio y ruido blanco gaussiano aditivo. Pero esta descripción se adapta a muchos sistemas prácticos de comunicación en un grado razonable, por lo que sistema hipotético ideal entrega información a una velocidad $\mathscr{C} = B \log(1 + S/N)$ es lo normal que por lo general se acepta para comparaciones de sistemas. En esta sección se reexaminan los diferentes sistemas existentes a la luz de la teoría de la información y se ve cómo se comportan éstos respecto del ideal.

Un aspecto de particular interés es la reducción de ruido de banda ancha. Por lo que, como preliminar, se investiga el intercambio de ancho de banda por relación señal a ruido (o potencia transmitida) que implica la ley de Hartley-Shannon, dado que éste es el intercambio ancho de banda-potencia óptima.

Intercambio ancho de banda-potencia óptima

Supóngase que se desea transmitir una señal, de banda limitada en W, tal que la relación señal a ruido de salida en el destino sea $(S/N)_D$. Sin importar como se efectúe la transmisión, la velocidad de la información en la salida no puede ser mayor que $\mathscr{R}_{\text{máx}} = W \log[1 + (S/N)_D]$. Supóngase además que se dispone de un sistema ideal para este objetivo y que el ancho de banda del canal (de transmisión) es B_T, la densidad de potencia de ruido es η, y la potencia promedio de la señal en el receptor es S_R. En otras palabras, la capacidad del canal es $\mathscr{C} = B_T \log[1 + (S/N)_R]$, en donde $(S/N)_R = S_R/\eta B_T$. En la figura 9.6 se resumen estos factores.

Si la información no se destruye ni se acumula en el receptor, la velocidad de salida es igual a la velocidad de información en el canal. Soponiendo que el sistema esta oprando a su capacidad y que la velocidad de salida es máxima, entonces $\mathscr{R}_{\text{máx}} = \mathscr{C}$, por lo que

$$B_T \log\left[1 + \left(\frac{S}{N}\right)_R\right] = W \log\left[1 + \left(\frac{S}{N}\right)_D\right]$$

Resolviendo para $\left(\frac{S}{N}\right)_D$ se tiene

$$\left(\frac{S}{N}\right)_D = \left[1 + \left(\frac{S}{N}\right)_R\right]^{B_T/W} - 1 \qquad (1)$$

lo cual demuestra que el intercambio óptimo de ancho de banda por potencia es *exponencial*. Para recalcar esta relación, nótese que $(S/N)_D \approx (S/N)_R^{B_T/W}$ sin las relaciones señal a ruido son grandes. El intercambio exponencial se efectúa por medio de un sistema ideal operando a su máxima capacidad.

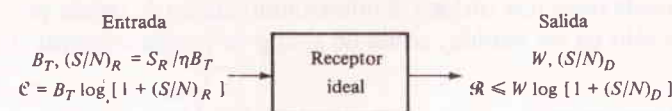



FIGURA 9.6

Sin embargo, con una *densidad de ruido* de canal fija, para diferenciarla de la potencia fija de ruido, la ecuación (1) no expresa lo mismo. Porque conforme el ancho de banda del canal aumenta, la potencia de ruido $N_R = \eta B_T$ se incrementa también; y $(S/N)_R$ disminuye. Se obtiene una base más equitativa para la comparación, reescribiendo la ecuación (1) en términos de los parámetros normalizados

$$\gamma = \frac{S_R}{\eta W} \qquad \text{y} \qquad \mathscr{B} = \frac{B_T}{W}$$

que se emplearon en capítulos anteriores.

La relación señal a ruido en la entrada del receptor es entonces $(S/N)_R = (SR/\eta W/)(W/B_T) = \gamma/\mathscr{B}$, y la ecuación (1) resulta ser

$$\left(\frac{S}{N}\right)_D = \left(1 + \frac{\gamma}{\mathscr{B}}\right)^{\mathscr{B}} - 1 \qquad (2)$$

$$\approx \left(\frac{\gamma}{\mathscr{B}}\right)^{\mathscr{B}} \qquad \frac{\gamma}{\mathscr{B}} \gg 1$$

Así, mientras que el intercambio no es exponencial en forma estricta, sí resulta muy cercano a ello para relaciones señal a ruido grandes. Esto significa que duplicando el ancho de banda de transmisión de un sistema ideal se eleva al cuadrado (en forma aproximada) la relación señal a ruido de salida. De manera alterna, puesto que γ es proporcional a S_T, se puede reducir la potencia transmitida a aproximadamente la raíz cuadrada de su valor original sin reducir a $(S/N)_D$ si se incrementa el ancho de banda en un factor de 2. Como recién se demostró, este intercambio resulta mucho mejor que el de muchos sistemas existentes.

La ecuación (2) demuestra también lo que está involucrado en la *compresión* del ancho de banda —la transmisión de una señal de ancho de banda W en un canal de ancho de banda $B_T < W$, de tal manera que $\mathscr{B} < 1$. Dando valores típicos, se encuentra que tal compresión es excesivamente costosa en términos de la potencia transmitida. Como aclaración, supóngase que se desea $(S/N)_D = 10^4$ y supóngase que $\eta W = 10^{-6}$. La transmisión en banda base ($\mathscr{B} = 1$) requiere $S_R = \eta W(S/N)_D =$ 10 mW. Pero para comprimir el ancho de banda en un factor de ½ se necesita que $\gamma = \mathscr{B}[1 + (S/N)_D]^{1/\mathscr{B}} - \mathscr{B} \approx \frac{1}{2}(10^4)^2 = 5 \times 10^7$, o $S_R = 50$ W, la cual es 5 000 veces la potencia de banda base. ¡De manera similar, para $\mathscr{B} = 1/_{10}$, el requisito de potencia es un colosal 10^{33} W!

Como corolario se puede decir que incluso el intercambio ancho de banda potencia óptimo, es práctico sólo en un sentido, aquel de ancho de banda creciente y potencia decreciente.

Transmisión de una señal analógica

Es algo difícil evaluar la velocidad de información de señales analógicas: formas de onda de voz y música, video de televisión, etc. Además, los sistemas de comunicación diseñados para tales señales tienen como objetivo la reproducción razonablemente fiel de las propias señales, con un mínimo de ruido y distorsión. Este objetivo no es exactamente lo mismo que la transferencia de información segura en el sentido de la teoría de la información; es decir, que al ingeniero de comunicaciones debe interesarle más el ancho de banda de transmisión, los requisitos de potencia de umbral y las relaciones señal a ruido que la capacidad del canal y su utilización.

No obstante, la teoría de la información sí tiene algo que ver en relación con la transmisión de señales analógicas. De manera específica, expresa la mejor relación señal a ruido que se puede obtener con unos parámetros de canal dados; dice la mínima potencia que se requiere para lograr una relación señal a ruido específica, como función del ancho de banda; e indica el intercambio óptimo posible de ancho de banda por potencia. Por tanto, compárese el comportamiento de los sistemas existentes con el de un sistema ideal como el descrito por la ecuacion (2)

En la tabla 9.1 se resumen muchos de estos resultados. Se supone que las relaciones señal a ruido son grandes, que todos los sistemas están encima del umbral, y el mensaje es normalizado de manera que $x^2 = ½$. Los valores para los sistemas pulsados suponen que $f_s = 2W$, etc., y no se lograría en la práctica. Es obvio que ninguno de los sistemas prácticos presentan la mejoría en la salida que se puede obtener en un sistema ideal incrementado *a uno u otro*: a γ (potencia) o a $\mathscr{B}$ (ancho de banda). La modulación por pulsos codificados sí tiene una dependencia exponencial de ancho de banda, pero, una vez encima del umbral, la potencia transmitida en aumento no produce una mejora adicional de

Tabla 9.1 COMPARACIÓN DE SISTEMAS DE TRANSMISIÓN ANALÓGICA DE MENSAJES

Sistema	$(S/N)_D$
AM	$\gamma/3$
SSB	γ
PDM	$\mathscr{B}\gamma/4$
PPM	$\mathscr{B}^2\gamma/16$
WBFM	$3\mathscr{B}^2\gamma/8$
DM†	$3\mathscr{B}^3/\pi^2$
PCM	$3\mu^2\mathscr{B}/2$
Ideal	$(\gamma/\mathscr{B})^{\mathscr{B}}$

† Suponiendo que $f_0 = W$.

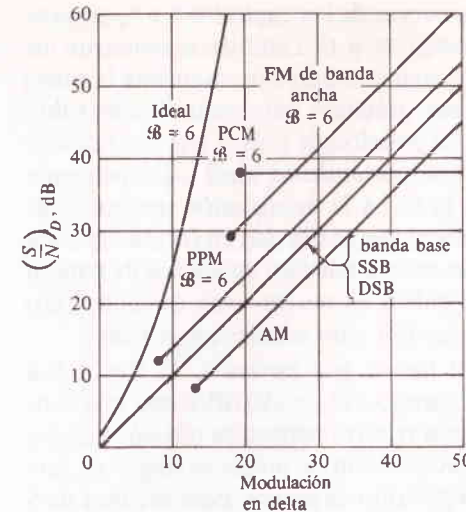



FIGURA 9.7
S/N de postdetección contra $\gamma = S_R/\eta W$.

$(S/N)_D$; su valor se determina por medio de la cuantificación. También se podría advertir que la banda lateral única (SSB) es casi tan buena como un sistema ideal que tenga $\mathscr{B} = 1$. Por supuesto que la relación de anchos de banda en banda lateral única está fija en $\mathscr{B} = 1$, por lo que no hay posibilidad de reducción de ruido de banda ancha.

Puesto que, en general, la relación señal a ruido de salida depende de γ como de $\mathscr{B}$, resulta difícil proporcionar un desplegado gráfico completo de las relaciones en la tabla 9.1. Como alternativa, se puede construir de gráfica $(S/N)_D$ como una función de γ para relaciones típicas de anchos de banda (Fig. 9.7), o del valor de γ que se requiere para una $(S/N)_D$ específica como una función de $\mathscr{B}$ (Fig. 9.8).

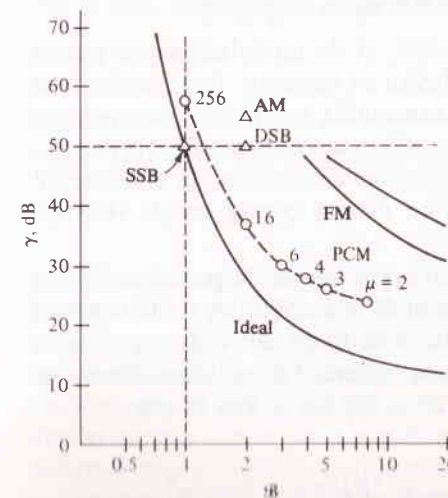



FIGURA 9.8
Potencia normalizada $\gamma = S_R/\eta W$ que se requiere para $(S/N)_D = 50$ dB como una función de la relación de anchos de banda $\mathscr{B} = B_T/W$.

En la figura 9.7 se repiten algunas de las curvas de los capítulos 7 y 8, aumentadas en una curva para un sistema ideal que tenga $\mathscr{B} = 6$. También se muestran los puntos de umbral de los sistemas prácticos —lo cual constituye precisamente la razón de ser de la figura. Se puede ver que los sistemas prácticos para reducción de ruido de banda ancha, como la FM, la modulación en posición de pulsos y la modulación por codificación de pulsos, quedan lejos del comportamiento ideal principalmente por las limitaciones de umbral. Por ejemplo, la FM y la modulación por codificación de pulsos con $\mathscr{B} = 6$ tienen puntos de umbral que se desvían en forma horizontal alrededor de 8 dB de un sistema ideal con la misma relación de anchos de banda. El umbral en la modulación en posición de pulsos es mucho más cercano, pero ocurre muy abajo de un valor de $(S/N)_D$ que sea útil para señales analógicas.

En cuanto al intercambio de ancho de banda por potencia, la figura 9.8 muestra el valor mínimo de γ que se necesita para $(S/N)_D = 50$ dB como una función de la relación de anchos de banda. Para esta relativamente alta relación señal a ruido de salida, se ve que la modulación por codificación de pulsos es mejor en forma considerable que la FM o la modulación en posición de pulsos, pero requiere de 6 a 8 dB más potencia que un sistema ideal. (Cómo se obtiene la curva para modulación por codificación de pulsos se examina a partir del siguiente encabezado.) Se podría puntualizar también el incremento agudo en γ para el sistema ideal cuando $\mathscr{B} < 1$, repitiendo la reciente observación respecto de la compresión del ancho de banda.

En resumen; a altas relaciones señal a ruido, la FM y la modulación por codificación por pulsos dan el mejor comportamiento en banda ancha, siendo algo mejor la segunda. Desde el punto de vista potencia-ancho de banda, todos los sistemas prácticos de banda ancha están en orden de magnitud abajo del ideal. A bajas relaciones señal a ruido, sólo la banda lateral única y la doble banda lateral son útiles, en las que no se presenta el efecto de umbral.

La capacidad de canal de la modulación por codificación de pulsos

De los diferentes sistemas que se han examinado, el de modulación por pulsos codificados es el más tratable para el análisis directo en términos de la teoría de la información. Esto sucede así porque la señal transmitida es discreta, aunque éste representa una señal analógica, y la velocidad de información se puede calcular con facilidad. Por lo tanto, se debe encontrar la capacidad del canal en la modulación por pulsos codificados y hacer una comparación directa con la ley de Hartley-Shannon.

Considérese un sistema de modulación por codificación de pulsos de banda base que tenga un ancho de banda de transmisión B, μ amplitudes codificadas de pulso espaciadas en forma igual, y relación señal a ruido de canal $(S/N)_R$. Puesto que la entropía de la señal digital es $\mathscr{H} \leq \log \mu$ y la velocidad de señalización es $r \leq 2B$, la velocidad de información en el canal es $\mathscr{R} \leq 2B \log \mu$. Por lo que

$$\mathscr{C} = \mathscr{R}_{\text{máx}} = 2B \log \mu = B \log \mu^2$$

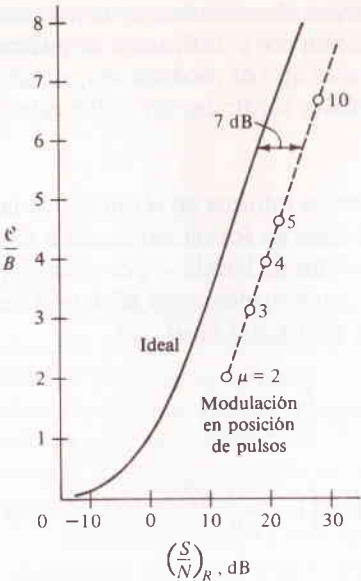



FIGURA 9.9
Capacidad de canal por ancho de banda unitario para modulación por pulsos codificados (con $P_e \sim 10^{-4}$) en comparación con un sistema ideal.

siempre y cuando se puedan despreciar los errores de decodificación, pero se determinó antes que esta clase de errores son despreciables si el sistema esta arriba del umbral, lo cual requiere que

$$\mu^2 \leq 1 + \frac{1}{5}\left(\frac{S}{N}\right)_R \tag{3}$$

Por lo que, exactamente arriba del umbral,

$$\mathscr{C} = B \log \left[1 + \frac{1}{5}\left(\frac{S}{N}\right)_R\right] \tag{4}$$

por lo que, si $(S/N)_R \gg 5$, $\mathscr{C} \approx B \log [(S/N)_R/5]$ o

$$\mathscr{C} \approx \mathscr{C}_{\text{ideal}} - B \log_2 5 \tag{5}$$

Basada en la ecuación (4), la figura 9.9 es la gráfica de $\mathscr{C}/B$ contra $(S/N)_R$. También se da para comparación la curva correspondiente para un sistema ideal. Visto de esta manera, se aprecia que la modulación por codificación de pulsos requiere alrededor de 7 dB más potencia que un sistema ideal. Sin embargo, es bueno llevar en mente que un sistema ideal tendría probabilidades de error pequeñas tendientes a desaparecer, mientras que la curva de modulación por codificación por pulsos es para $P_e \approx 10^{-4}$. La razón por la que la modulación por codificación de pulsos se compara tan favorablemente con un sistema ideal, procede de la reciente

conclusión de que, dado el ruido del canal, la comunicación eléctrica es de manera inherente un proceso discreto. El diseño de la modulación por codificación de pulsos reconoce y acepta este hecho; la señal transmitida de este tipo de modulación, siendo discreta, se adapta mejor al canal ruidoso que las señales continuas no codificadas.

Eficiencia en la comunicación

Por último, se deben examinar los requisitos de potencia mínima en términos de la velocidad de información. Este punto de vista es aplicable en forma particular a sistemas de largo alcance que no tienen restricción de ancho de banda —por ejemplo, los sistemas de comunicación espacial— y conduce a la tan mencionada eficiencia de comunicación. Como en lo anterior, la referencia es el sistema ideal con

$$\mathscr{C} = B_T \log\left(1 + \frac{S_R}{\eta B_T}\right)$$

$$= \frac{S_R}{\eta} \log\left(1 + \frac{S_R}{\eta B_T}\right)^{\eta B_T/S_R} \tag{6}$$

lo cual se ha reescrito para demostrar que si S_R y η son fijos, $\mathscr{C}$ se aumenta al máximo haciendo que $B_T \to \infty$. Por tanto, aprovechando que $\lim_{\nu \to 0}(1 + \nu)^{1/\nu} = e$, la velocidad máxima de información en un sistema ideal es

$$\mathscr{R}_{\text{máx}} = \lim_{B_T \to \infty} \mathscr{C} = \frac{S_R}{\eta} \log_2 e = 1.44 \frac{S_R}{\eta} \tag{7a}$$

o, para una velocidad específica de información $\mathscr{R}$,

$$S_{R\,\text{mín}} = 0.693\, \eta \mathscr{R} \tag{7b}$$

donde ambas requieren una técnica de codificación tal que el ancho de banda de transmisión se acerque al infinito. ¡Y al mismo tiempo $(S/N)_R = S_R/\eta B_T$ tiende a cero!

Ahora, considérese cualquier sistema que tenga una velocidad de información y una potencia recibida S_R. Su eficiencia de comunicación se puede definir como $\mathscr{E} = \mathscr{R}/\mathscr{R}_{\text{máx}} = S_{R\text{mín}}/S_R$ o, empleando la ecuación (7)

$$\mathscr{E} \triangleq 0.693 \frac{\eta \mathscr{R}}{S_R} \tag{8}$$

por lo que si $\mathscr{E} = 0.2$, por ejemplo, el sistema requiere cinco veces más potencia que un sistema ideal operando con $B_T \to \infty$. Para subrayar el significado de $\mathscr{E}$, hágase $\eta = k\mathscr{T}_N$ donde k es la constante de Boltzmann y $\mathscr{T}_N$ la temperatura de ruido del sistema; si $\mathscr{L}$ es la pérdida de transmisión, entonces la potencia *transmitida* que se requiere es

$$S_T = \mathscr{L} S_R = 0.693 k \frac{\mathscr{L}\mathscr{T}_N \mathscr{R}}{\mathscr{E}}$$

$$\approx 10^{-23} \frac{\mathscr{L}\mathscr{T}_N \mathscr{R}}{\mathscr{E}} \tag{9}$$

Es obvio que reduciendo al mínimo a S_T da por resultado el aumento al máximo de $\mathscr{E}$.

En relación con el dibujo de los sistemas prácticos eficientes, se infiere de la ecuación (7) que se aumenta al máximo a $\mathscr{E}$ si se hace a B_T tan grande como sea posible y a $(S/N)_R$ tan pequeño como se pueda. Se puede obtener un B_T grande con técnicas de *modulación de banda ancha*, pero el inevitable efecto de umbral prohíbe valores muy pequeños de $(S/N)_R$ y con ello se impide la operación a la mayor eficiencia. Este punto se demuestra con mayor amplitud examinando de nuevo la modulación por pulsos codificados.

Suponiendo que la velocidad de información en la modulación por pulsos codificados es $\mathscr{R} = \mathscr{C}$, con $\mathscr{C}$ dada por la ecuación (4), la eficiencia es

$$\mathscr{E} = 0.693 \frac{\eta}{S_R} B_T \log\left[1 + \frac{1}{5}\left(\frac{S}{N_R}\right)\right]$$

$$= \frac{0.693}{(S/N)_R} \log\left[1 + \frac{1}{5}\left(\frac{S}{N}\right)_R\right]$$

Sin embargo, de la ecuación (3), la condición de umbral es $(S/N)_R > 5(\mu^2 - 1)$, por lo que no se puede tomar a $(S/N)_R$ arbitrariamente pequeña. Lo mejor que se puede hacer es $\mu = 2$ (modulación por pulsos codificados *binaria*), lo cual da el mayor ancho de banda de transmisión, la menor relación señal a ruido del canal y la mayor eficiencia. Con anterioridad, varias veces se ha sospechado que la modulación por pulsos codificados binaria es superior a la modulación por pulsos codificados con $\mu > 2$; esta sospecha se confirma ahora desde el punto de vista de la potencia. Para $\mu = 2$, $(S/N)_R \geq 15$, y

$$\mathscr{E} = \frac{0.693}{15} \log 4 = 0.0924$$

Por tanto, la eficiencia está en la vecindad del 9%. Esta baja eficiencia puede parecer desalentadora al principio. Situando las cosas en la perspectiva apropiada, es mejor decir que la modulación por codificación de pulsos binaria requiere alrededor de 10 dB más potencia que un sistema *ideal* con ancho de banda *infinito*.

Volviendo a la modulación analógica, se enfrenta con el problema de la estimación de la velocidad de información $\mathscr{R}$ de una señal analógica.† Un recurso imperfecto pero sencillo es tomar el límite superior $\mathscr{R} \leq W \log[1 + (S/N)_D]$, donde

† Sanders (1960) proporciona un análisis más completo.

$(S/N)_D$ y W son los parámetros de la señal después de la demodulación. Esto proporciona un límite superior en la eficiencia, a saber,

$$\mathscr{E} \leq \frac{0.693}{\gamma} \log \left[1 + \left(\frac{S}{N}\right)_D\right] \qquad (10)$$

Para modulación lineal de portadora suprimida (banda lateral única y doble banda lateral), se tiene $(S/N)_D = \gamma$ y tales sistemas se operan en forma normal con relaciones señal a ruido de 30 a 40 dB. La eficiencia correspondiente es sustancialmente menor del 1%, lo que no se esperaba, puesto que el ancho de banda de transmisión es relativamente pequeño. Por otra parte, la modulación analógica de banda ancha debe demostrar que es mejor, y de hecho lo es.

Para FM (sin desacentuación) se puede demostrar que hay una relación de desviación óptima, $\Delta_{\text{ópt}} \approx 2$, dando $\mathscr{E} < 0.1$; con $\Delta < 2$ la pérdida de reducción de ruido de banda ancha disminuye la eficiencia, mientras que para $\Delta > 2$ el nivel de umbral superior equilibra la mejora de S/N y de nuevo disminuye la eficiencia. Es obvio que las técnicas de extensión de umbral mencionadas en la sección 7.5 serían benéficas aquí, puesto que ellas disminuyen el valor γ-ésimo. En teoría, la FM con realimentación compresiva de frecuencia en el receptor (FMFB) es capaz de lograr el 50% del máximo de eficiencia; en la práctica, son posibles las eficiencias de 25 al 30%. La modulación en posición de pulsos parece tener eficiencias comparables a la FMFB gracias a su bajo nivel de umbral.

Ejemplo 9.4

Para aclarar las implicaciones de la ecuación (9), supóngase que se desea transmitir una imagen fija de Marte a la Tierra con $S_T = 10$ Wz, $\mathscr{L} = 200$ dB y $\mathscr{T}_N = 50°k$. Si el sistema tiene 100% de eficiencia ($\mathscr{E} = 1$), la velocidad máxima de información es

$$\mathscr{R} = 10^{23} \frac{\mathscr{E} S_T}{\mathscr{L}\mathscr{T}_N} = 200 \text{ bits/s}$$

Suponiendo que la fotografía se cuantifica en 200 × 200 = 4 × 10⁴ elementos, teniendo cada elemento uno de los 64 niveles posibles de brillantez, la información total a transferirse es $\mathscr{I} = 24 \times 10^4$ bits (véase el ejemplo 9.1). Por lo tanto, el tiempo total de transmisión T debe ser

$$T = \frac{\mathscr{I}}{\mathscr{R}} = \frac{24 \times 10^4}{200} = 1{,}200 \text{ s} = 20 \text{ min}$$

La misión Mariner IV, empleando un sistema menos eficiente, pero práctico, necesitó 8 h para transmitir cada imagen.

9.5 PROBLEMAS

9.1 (Secc. 9.1) Enúnciense en palabras e interprétense los requisitos para $f(\ )$ en las ecuaciones de la (1) a la (3).

9.2 (Secc. 9.1) Se selecciona una carta en forma aleatoria de una baraja de 52. Se le dice a usted que se trata de un corazón. ¿Cuánta información (en bits) ha recibido usted? ¿Cuánta información más se necesita para especificar en forma completa la carta? *Respuesta*: 2 bits, 3.7 bits.

9.3 (Secc. 9.1) Calcúlese la cantidad de información que se necesita para abrir un candado cuya combinación son tres números, abarcando cada uno de 00 a 99.

9.4 (Secc. 9.1) Una fuente produce seis símbolos con probabilidades $1/2$, $1/4$, $1/8$, $1/16$, $1/32$ y $1/32$. Encuéntrese la entropía $\mathscr{H}$.

9.5 (Secc. 9.1) Una fuente tiene un alfabeto de tamaño m. Un símbolo tiene una probabilidad ϵ mientras que los otros son igualmente probables. Encuéntrese $\mathscr{H}$ en términos de m y de ϵ.

9.6★ (Secc. 9.1) Demuéstrese que $\Sigma_{j=1}^{m} P_j \log_2(1/mP_j) = \mathscr{H} - \log_2 m$ y empléese esto con el hecho de que $\ln v \leq v - 1$, para demostrar que $\mathscr{H} \leq \log_2 m$.

9.7★ (Secc. 9.1) Supóngase que se dan nueve centavos, ocho de los cuales son buenos. La moneda restante es falsa y pesa más o menos lo que una buena. Dada una balanza sin calibrar, ¿cuántas veces se necesita pesar en teoría para localizar la moneda mala y determinar si es ligera o pesada? (*Sugerencia*: La balanza tiene tres posiciones posibles, balanceada y desbalanceada sobre un lado o sobre el otro. La información promedio por pesada, es decir, la entropía, se aumenta al máximo si cada una de estas posiciones es igualmente probable. Aunque no es necesario, quizá se desee idear el procedimiento correcto de pasaje.)

9.8 (Secc. 9.1) Una fuente de datos determinada tiene ocho símbolos que se producen en bloques de tres a una velocidad de 1 000 bloques por segundo. El primer símbolo en cada bloque es siempre el mismo (se presume que para sincronización); los dos lugares restantes se llenan con cualquiera de los ocho símbolos con igual probabilidad. ¿Cuál es la velocidad de la entropía $\mathscr{R}$? *Respuesta*: 6 000 bits/s.

9.9 (Secc. 9.1) Una fuente de datos determinada tiene 16 símbolos posibles y equiprobables, cada uno de 1 milisegundo (ms) de longitud. Los símbolos se envían en bloques de 15, separados por un pulso de sincronización de 5 ms. Calcúlese $\mathscr{R}$.

9.10 (Secc. 9.1) Una fuente de datos binarios tiene $P_0 = 3/8$, $P_1 = 5/8$, e influencia intersimbólica que se extiende sobre grupos de dos símbolos sucesivos tal que $P(1/0) = 3/4$ y $P(0/1) = 1/16$. Calcúlese la entropía condicional $\mathscr{H}_c$ y compárese con $\mathscr{H}$.

9.11 (Secc. 9.1) En forma análoga a la ecuación (7), la entropía de una *señal continua*, se puede definir como

$$\mathscr{H}(x) = -\int_{-\infty}^{\infty} p(x) \log p(x)\, dx$$

donde $p(x)$ es la función de densidad de probabilidad. Demuéstrese que esto es más una medida relativa que absoluta de información, considerando a $\mathscr{H}(x)$ y a $\mathscr{H}(y)$ cuando $y(t) = Kx(t)$ y $x(t)$ está uniformemente distribuida sobre $[-1, 1]$.

9.12 (Secc. 9.2) Verifíquese que $P(1) = P(0)$ en el ejemplo 9.3.

9.13 (Secc. 9.2) Un canal discreto sin ruido va a conducir información a una velocidad de 900 bits/s. Determínese el número mínimo de estados de canal que se requiere si $r = 200$ o 1 000. *Respuesta*: 23, 2.

9.14 (Secc. 9.2) Una fuente discreta produce los símbolos A y B con $P_A = 3/4$ y $P_B = 1/4$ a una velocidad de 100 símbolos/s. En un intento de acoplar la fuente a un canal binario sin ruido, los símbolos se agrupan en bloques de dos y se codifican como sigue:

Símbolos agrupados	Código binario
AA	1
AB	01
BA	001
BB	000

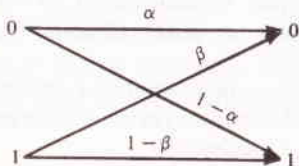
Calculando la velocidad de entropía de la fuente y la velocidad de símbolos del canal, demuéstrese que este código no es óptimo pero es razonablemente eficiente.

9.15★ (Secc. 9.2) Considérese un canal binario con ruido que afecta a los símbolos de canal en bloques de tres, tales que se recibe un bloque sin errores o que hay exactamente un error en el primero, segundo o tercer dígito. Estas cuatro posibilidades son equiprobables.

(*a*) Demuéstrese que $\mathscr{C} = r/3$.

(*b*) Idéese un código que produzca una transmisión exenta de error en $\mathscr{R} = \mathscr{C}$.

9.16 (Secc. 9.2) En la figura P9.1 se indican en forma esquemática las probabilidades de error por dígito y de no error para un *canal binario no simétrico*, es decir, P(0 recibido 0 enviado) $= \alpha$, etc. Demuéstrese que

$$\mathscr{R}_{cc} = r\left(\frac{1-\alpha+\beta}{2}\log\frac{2}{1-\alpha+\beta} + \frac{1+\alpha-\beta}{2}\log\frac{2}{1+\alpha-\beta}\right)$$

9.17 (Secc. 9.2) Se dice del canal del problema 9.16 que es *inútil* si $\alpha = \beta$. Justifíquese esto en forma matemática e intuitiva.

9.18 (Secc. 9.3) Un canal discreto sin ruido tiene $r = 10^5$. Investíguese la posibilidad de reemplazarlo con un canal continuo que tenga $B = 8$ kHz y $S/N = 31$.

9.19 (Secc. 9.3) Encuéntrese el tiempo mínimo que se requiere para transmitir 600 dígitos *decimales* en el canal continuo del problema 9.18. *Respuesta*: 0.05 s.

9.20 (Secc. 9.3) Un sistema de comunicación determinado emplea pulsos de RF con cuatro niveles de amplitud posibles. Si se desea una velocidad de información de 10^6 bits/s, ¿cuál es el valor mínimo práctico para la frecuencia portadora f_c? (*Sugerencia*: Recuérdense las restricciones del ancho de banda fraccionario.)

9.21 (Secc. 9.3) Estímese a $M(T)$ para un canal telefónico típico con $B = 3$ kHz, $S/N = 30$ dB y $T = 3$ min.

9.22★ (Secc. 9.3) Considérese el conjunto de señales de canal $x_i(t) = \text{senc}\,(2Bt - i)$, $i = 0, \pm 1, \pm 2, \ldots$ Examínese la interpretación del espacio de la señal y determínese el límite superior en superior en i para un espacio de D dimensiones.

9.23★ (Secc. 9.3) Un canal continuo tiene $S/N = 63$ y $B = 1$ kHz. Obténganse límites numéricos en M y T tales que $\mathscr{R} = 5\,000$ con $P_e \leq 10^{-3} \approx +^{-10}$.

9.24 (Secc. 9.4) Un ingeniero pretende haber diseñado un sistema de comunicación que da $(S/N)_D = 10^6$ cuando $(S/N)_R = 3$ y $B_T = 10$ W. ¿Le cree usted? Explíquelo.

9.25 (Secc. 9.4) Un sistema ideal tiene $\mathscr{B} = 4$ y $(S/N)_D = 40$ dB. ¿Cuál es el nuevo valor de $(S/N)_D$ si B_T se triplica mientras todos los otros parámetros permanecen fijos? *Respuesta*: 72 dB.

9.26 (Secc. 9.4) Repítase el problema 9.25 con B_T cambiada a $W/2$ en vez de que se triplique.

9.27 (Secc. 9.4) Un sistema de comunicación tiene $B_T = 5$ kHz, $\mathscr{L} = 30$ dB, y $\eta = 10^{-7}$ W/Hz. Encuéntrese el mínimo valor teórico para S_T que produzca $(S/N)_D = 50$ dB cuando $W = 1$ kHz.

9.28★ (Secc. 9.4) Considérese la eficiencia de un sistema ideal con B_T finito.

(*a*) Demuéstrese que $\mathscr{E} = 0.693(\mathscr{R}/B_T)/(2^{\mathscr{R}/B_T} - 1)$.

(*b*) Bosquéjese a $\mathscr{E}$ contra $B_T/\mathscr{R}$ y examínense las implicaciones de esta curva.

9.29★ (Secc. 9.4) Empleando la ecuación (10) y la regla de Carson, obténgase un límite superior en $\mathscr{E}$ en términos de Δ para sistema de FM (sin desacentuación) operando justo por encima del umbral.

9.30★ (Secc. 9.4) Supóngase que se desea tener una transmisión de voz en vivo de Marte a la Tierra por medio de modulación por codificación de pulsos binaria con $f_s = 8$ kHz y $Q = 64$. Estímese el requisito de potencia en el transmisor de Marte empleando valores típicos en la ecuación (9).

10

SISTEMAS DIGITALES DE DATOS

Hasta alrededor de 1950, la transmisión analógica de señales era la única mercancía disponible de la industria de la comunicación, salvo, por supuesto, el telégrafo y el teletipo. Pero eso fue antes del advenimiento de la automatización y de la revolución causada por los computadores. Hoy, se generan enormes cantidades de datos digitales en el gobierno, el comercio y la ciencia. Además, la necesidad de eliminar el manejo manual y los errores humanos, condujo al concepto de los sistemas integrados para el procesamiento de datos. Estos desarrollos, con una tendencia creciente hacia la descentralización, hicieron de la transmisión y distribución de las señales digitales la tarea más vasta de la comunicación eléctrica.

Mientras que de acuerdo con la teoría de la información es posible la transmisión de datos con mucha eficiencia y virtualmente sin errores, la mayor parte de las aplicaciones sencillamente no justifica el costo y la complejidad que demanda un sistema de comunicación cercano al ideal. Esto no quiere decir que la teoría de la información no tenga cabida en el estudio de la transmisión de datos, sino más bien que puede ser mejor hacer el trabajo ahora, en una forma menos que óptima, que esperar hasta que los avances tecnológicos permitan una solución más compleja. En consecuencia, el diseño de un sistema práctico de datos es por lo general bastante pragmático, basado en dos consideraciones elementales: la velocidad tosca de fuente en dígitos por segundo para diferenciarla de la velocidad de entropía en bits por segundo; y la seguridad buscada en la transmisión, es decir, la frecuencia tolerada de error. Antes de continuar, es quizá útil tener cierto tacto para los valores típicos de estos parámetros.

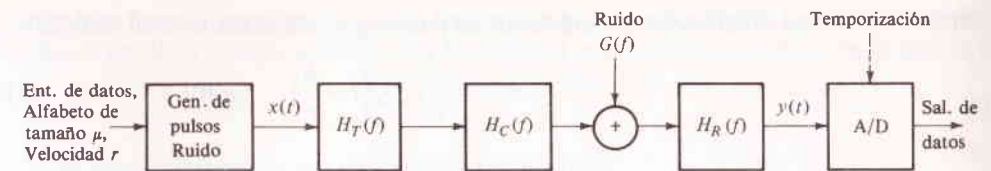



FIGURA 10.1
Sistema de datos en banda base

Así como las señales analógicas tienen un intervalo amplio de anchos de banda, también las señales digitales tienen un intervalo amplio de *velocidad de bits*,† dependiendo de la fuente y la aplicación. La salida de los dispositivos de teclado manual, como el teleimpresor, raras veces es mayor de 100 bits/s. Las velocidades en telemetría pueden ser aún menores. En el extremo opuesto, una señal de TV a color en modulación por codificación de pulsos requiere, a groso modo, unos 100 megabits/s. Se obtienen velocidades intermedias de 1 a 100 kilobits/s por medio de lectores de datos electromecánicos y electromagnéticos. Es común que los mejores sistemas de banda ancha tengan capacidades de velocidad de varios cientos de megabits por segundo. Pero la mayor parte de la transmisión digital se efectúa por medio de circuitos telefónicos de voz normales con velocidades límite de 1 000 a 5 000 bits/s.

A diferencia de la rapidez de señalización, la cuestión acerca de la frecuencia aceptable de error resulta bastante confusa. Es obvio que los errores son de grave implicación en los sistemas de agrupación de computadores (en especial, si se ve afectado en forma adversa el equilibrio del grupo de usted) mientras que los requisitos pueden ser menos exigentes si se están procesando datos en telemetría con instrumentos de medición de un 10% de precisión. Para lo primero es probable que sea intolerable una probabilidad de error de 10^{-5}; para lo último, una probabilidad de 10^{-3} sería un desperdicio rayando en la extravagancia. No obstante, las probabilidades de error del orden de 10^{-4} son representativas y adecuadas a muchas aplicaciones. Se tomará este valor como guía para propósitos de comparación, sin olvidar su naturaleza arbitraria.

En este capítulo serán examinados los sistemas digitales de datos, sus problemas y principios de diseño, a partir de donde se dejó, sección 4.5, con las técnicas de diseño en banda base, muchas de las cuales se aplican después a los sistemas de modulación digital. El capítulo se cierra con una introducción a la codificación para control de error.

10.1 DISEÑO DEL SISTEMA DE BANDA BASE

Los elementos funcionales de un sistema de datos en banda base se despliegan en la figura 10.1. Una fuente de datos que tenga un alfabeto de μ dígitos posibles (o

† Por convención, se emplean en este capítulo los términos *bit* y dígito binario en forma permutable.

símbolos) entrega r dígitos/s a un generador de pulsos que conforma la señal analógica

$$x(t) = \sum_k a_k p_S\left(t - \frac{k}{r}\right) \qquad p_S(0) = 1 \tag{1}$$

donde $p_s(t)$ es la forma de pulso normalizado y los a_k representan a los dígitos de fuente. Después de pasar a través de un filtro de transmisión $H_T(f)$ y del canal $H_c(f)$, la señal se contamina con el ruido aditivo, por lo que la señal más el ruido en la salida del filtro de recepción $H_R(f)$ es

$$y(t) = \sum_k A_k p\left(t - t_d - \frac{k}{r}\right) + n(t) \tag{2}$$

donde t_d es el retardo total en tiempo y K_R la amplificación, es decir,

$$A_k = K_R a_k \qquad p(0) = 1$$

El convertidor A/D opera sobre $y(t)$ muestreándola a los tiempos óptimos

$$t_m = \frac{m}{r} + t_d.$$

dando

$$y(t_m) = A_m + \sum_{k \neq m} A_k p\left(\frac{m-k}{r}\right) + n(t_m) \tag{3}$$

en la cual el primer término representa al dígito m-ésimo, mientras que los otros dos términos son *interferencia intersimbólica* y *ruido*, respectivamente. Por último, el mensaje digital (más errores) se regenera comparando a $y(t_m)$ con un conjunto de niveles de umbral.

La ecuación (3) muestra dos aspectos de diseño del sistema de banda base para una transmisión segura de datos, a saber, reduciendo el ruido y eliminando o reduciendo la interferencia intersimbólica. De manera adicional, si existe una restricción de ancho de banda, se puede desear aumentar al máximo la velocidad de señalización r para un ancho de banda de canal dado B o reducir al mínimo el ancho de banda que se requiere para una velocidad dada. Aquí se consideran estos aspectos. A menos que se establezca otra cosa, se presume que el ruido es gaussiano de valor medio cero con densidad espectral $G(f) = \eta/2$. La fuente se considera siempre como ergódica con dígitos equiprobables.

Sistema de banda base ideal

En la sección 4.5 se supuso que el canal era sin distorsión, con ancho de banda B, es decir, $H_c(f) = K_R e^{-j\omega t_d}$ cuando se incluye el retardo en tiempo. En conformidad, se tomó la forma de pulso de banda limitada $p_s(t) = \text{senc}\, rt$ con $r = B/2$ y siendo $H_R(f)$ un filtro pasabajas ideal. Por lo tanto, la velocidad de señalización es máxima

$$r = 2B \tag{4}$$

y no hay interferencia intersimbólica, puesto que

$$p(t) = \text{senc}\, rt \qquad p\left(\frac{m-k}{r}\right) = \begin{cases} 1 & m = k \\ 0 & m \neq k \end{cases} \tag{5}$$

También se supuso una señal polar con espaciamiento de nivel $2A$, es decir,

$$A_k = \begin{cases} 0, \pm 2A, \pm 4A, \ldots, \pm(\mu - 1)A & \mu \text{ impar} \\ \pm A, \pm 3A, \ldots, \pm(\mu - 1)A & \mu \text{ par} \end{cases} \tag{6}$$

por lo que la potencia de la señal recibida es

$$S_R = \overline{A^2} = \frac{(\mu^2 - 1)A^2}{3} \tag{7}$$

Se demostró después que, con ruido gaussiano de valor medio cero, los niveles de umbral óptimos están a la mitad, entre los valores de A_K y

$$P_e = 2\left(1 - \frac{1}{\mu}\right) Q\left(\frac{A}{\sigma}\right) \tag{8}$$

en donde σ es el valor rms de n por lo que

$$\sigma^2 = \overline{n^2} = N = \eta B = \frac{\eta r}{2}$$

Así

$$\left(\frac{A}{\sigma}\right)^2 = \frac{3}{\mu^2 - 1} \frac{S_R}{\eta B} = \frac{3}{\mu^2 - 1} \frac{S}{N} = \frac{6}{\mu^2 - 1} \rho \tag{9}$$

en donde

$$\rho \triangleq \frac{S_R}{\eta r} \tag{10}$$

Este sistema recibirá el nombre de sistema de banda base *ideal*, y su comportamiento sirve como un dato para comparaciones de sistemas —en especial en términos del parámetro normalizado ρ.

Sistema de pulsos rectangulares — filtraje de integrado y descarga

Gracias a su forma de pulso muy especial, el sistema ideal obtiene la máxima rapidez de señalización en un canal de banda limitada. Pero si no existe restricción de ancho de

banda, en la misma forma se podrían emplear pulsos rectangulares que no se traslapen

$$p_S(t) = \Pi\left(\frac{t}{\tau}\right) \qquad \tau \leq \frac{1}{r} \tag{11}$$

los cuales son mucho más sencillos de generar. Recordándo lo estudiado de las ondas binarias aleatorias (Ej. 3.7), (Sec. 3.5), se desprende que el ancho de banda disponible debe ser $B \gg 1/\tau \geq r$ para impedir que los pulsos se contaminen y evitar la influencia intersimbólica. Sin embargo, esto no significa que el filtro de recepción deba estar abierto de par en par para que se llegue a un excesivo ruido de salida y a altas frecuencias de error. Es claro que el elemento sobre el que gira tal sistema es $H_R(f)$, el cual se debe diseñar para que reduzca al mínimo en forma simultánea el ruido y la influencia intersimbólica debidos a la contaminación a la salida del filtro.

En cuanto al ruido, se sabe que un *filtro acoplado* tiene el efecto deseable de aumentar al máximo el pico de un pulso de salida en comparación con el ruido rms. En consecuencia, recurriendo a la ecuación (12*b*), (Sec. 4.4), supóngase que la respuesta al impulso del filtro de recepción es

$$h_R(t) = \Pi\left(\frac{t_0 - t}{\tau}\right) = \Pi\left(\frac{t - t_0}{\tau}\right) \tag{12}$$

conforme se bosquejó en la figura 10.2*a* para $t_0 = \tau/2$. Para analizar el comportamiento se supondrá una señal binaria y se despreciará cualquier distorsión o retardo

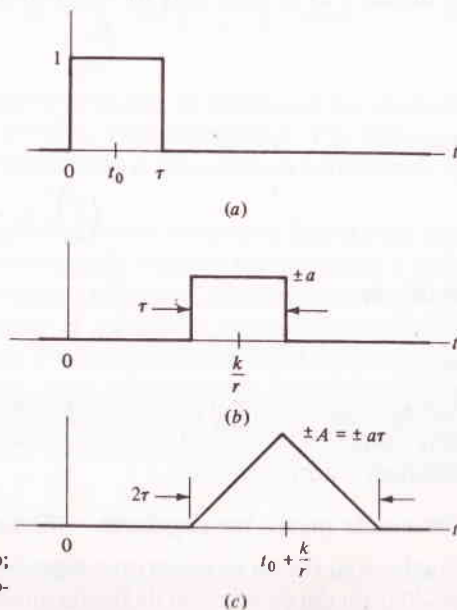



FIGURA 10.2
Filtraje acoplado. (*a*) Respuesta al impulso; (*b*) pulso recibido; (*c*) salida de filtro acoplado

en la transmisión, por lo que el pulso *k*-ésimo a la entrada del receptor es de la forma $\pm a\Pi[(t - k/r)/\tau]$, (Fig. 10.2*b*). Convolucionando con $h_R(t)$ se produce un pulso *triangular* de altura $\pm a\tau$ y anchura 2τ (Fig. 10.2*c*), es decir,

$$A_k p\left(t - \frac{k}{r}\right) = \pm A\Lambda\left(\frac{t - t_0 - k/r}{\tau}\right) \qquad A = a\tau \tag{13}$$

Es obvio que los tiempos óptimos de muestreo son ahora $t_m = (m/r) + t_0$ (en los cuales se debe agregar cualquier retardo en tiempo t_d), y *no existe interferencia intersimbólica*, puesto que $p(t - k/r) = 0$ para $|t - k/r| \geq 1/r$ si $\tau \leq 1/r$.

Suponiendo el muestreo a los tiempos óptimos, la probabilidad de error es $p_e = Q(A\sigma)$ para el caso binario sujeto a examen. Pero de la ecuación (12*a*), (Sec. 4.4) $(A/\sigma)^2 = 2E_R/\eta$ en donde E_R es la energía en el pulso recibido. Con pulsos binarios polares rectangulares, $E_R = (\pm a)^2\tau = S_R/r$, y

$$\left(\frac{A}{\sigma}\right)^2_{\text{máx}} = \frac{2E_R}{\eta} = \frac{2S_R}{\eta r} = 2\rho \tag{14}$$

expresión idéntica a la ecuación (9) con $\mu = 2$. El análisis que es similar para el caso μ-ario demuestra que la ecuación (9) aún es válida, y se concluye que la frecuencia de error de un sistema de pulsos rectangulares es casi tan buena como en un sistema ideal, siempre y cuando el filtro de recepción se diseñe de conformidad con la ecuación (12).

Desafortunadamente, esa condición resulta ser un obstáculo, porque no se puede sintetizar una respuesta al impulso rectangular en forma perfecta, y todas las aproximaciones tienen colas que decaen, las cuales originan interferenca intersimbólica. Pero, por lo general, existe más de una forma de resolver un problema y en relación con la figura 10.2*c* se demuestra que la parte posterior del pulso triangular después de $t = t_0 + K/r$ es innecesaria. Por lo tanto, supóngase que simplemente se integra cada pulso entrante y que después se restablece o "descarga" el integrador y en forma inmediata al tiempo de muestreo. Este filtraje de *integrado y descarga* se ocupa del problema de la interferencia intersimbólica (como se ilustra en la Fig. 10.3*a*) y permite el mismo valor máximo de A/σ. La figura 10.3*b* es la realiza-

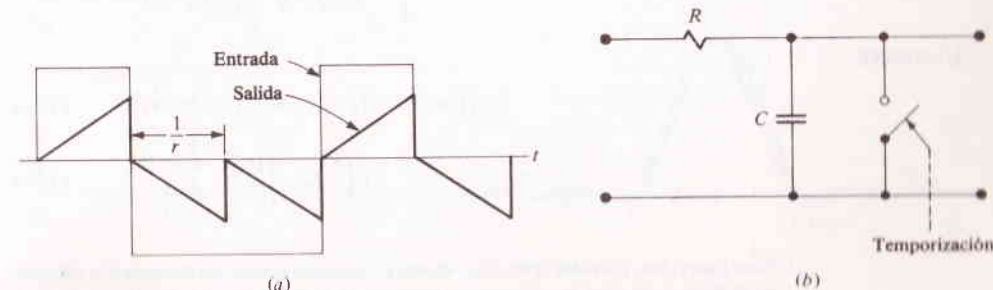



FIGURA 10.3
Filtraje de integrado y descarga. (*a*) Formas de onda; (*b*) circuito.

ción del circuito; si RC $\gg 1/r$, la salida aparece en forma lineal sobre la duración del pulso de entrada hasta que el cierre momentáneo del interruptor descarga al capacitor y restablece la salida en cero. No es necesario decir que el muestreo y la descarga se deben sincronizar con sumo cuidado. Aún así, el filtro de integrado y descarga representa de manera virtual el único caso donde el filtraje acoplado se acerca con bastante aproximación en la práctica.

EJERCICIO 10.1 Considérese un sistema de pulsos rectangulares con $\tau = 1/2r$ y un filtro de recepción que es un filtro pasabajas RC sencillo *sin* interruptor de descarga. Suponiendo que RC sea suficientemente pequeño como para que la interferencia intersimbólica sea despreciable, demuéstrese que

$$\left(\frac{A}{\sigma}\right)^2 = \frac{2S_R}{(\eta/4RC)} = 4RCr(2\rho)$$

Evalúese 4RC tal que los pulsos de salida decaigan a 0.001 de su valor máximo por el siguiente tiempo de muestra. *Respuesta*: 0.58.

Formación de pulsos de Nyquist

Entre los extremos de los pulsos senc ($B = r/2$) y los rectangulares ($B \gg r$), existe un compromiso que hace posible el empleo de los pulsos de formación de Nyquist, pulsos de banda limitada en $B \geq r/2$ que tienen cruzamientos cero periódicos para eliminar la interferencia intersimbólica. Se considerará sólo el caso de $r/2 \leq B \leq r$, el cual es del mayor interés práctico y para el cual se puede presentar el *teorema de la simetría residual* de Nyquist, como sigue:† Sea

$$P(f) = [P_\beta(f)] * \left[\left(\frac{1}{r}\right)\Pi\left(\frac{f}{r}\right)\right] \tag{15a}$$

con

$$P_\beta(f) = 0 \qquad |f| > \beta \leq \frac{r}{2}$$

$$\int_{-\infty}^{\infty} P_\beta(f)\, df = p_\beta(0) = 1 \tag{15b}$$

Entonces

$$p(t) = \mathscr{F}^{-1}[P(f)] = p_\beta(t)\,\text{senc}\, rt \tag{16a}$$

$$p\left(\frac{k}{r}\right) = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases} \tag{16b}$$

† Véase Lucky, Salz y Weldon (1968, Cap. 4) para el enunciado general y la demostración del criterio de Nyquist, así como la conformación especializada de pulsos conocida como *respuesta parcial* o duobinaria que acepta en intercambio cantidades controladas de interferencia intersimbólica para señalización más rápida.

y $P(f)$ es de banda limitada en $B = (r/2) + \beta \leq r$ por lo que

$$r = 2B - \beta$$

La demostración de estas aseveraciones debe ser autoevidente.

Un número infinito de funciones satisfacen las condiciones anteriores, incluyéndose el caso ideal $p(t)$ = senc rt; es decir, si $p_\beta(f) = \delta(f)$ entonces $p_\beta(t) = 1$, $\beta = 0$, y $r = 2B$. Una de estas funciones tiene la característica de frecuencia *de caída senoidal* o *de subida cosenoidal* como se ve en la gráfica de la figura 10.4*a*. En forma analítica,

$$P_\beta(f) = \frac{\pi}{4\beta}\cos\frac{\pi f}{2\beta}\,\Pi\left(\frac{f}{2\beta}\right) \qquad 0 < \beta \leq \frac{r}{2} \tag{17a}$$

por lo que

$$P(f) = \begin{cases} \dfrac{1}{r} & |f| < \dfrac{r}{2} - \beta \\ \dfrac{1}{r}\cos^2\dfrac{\pi}{4\beta}\left(|f| - \dfrac{r}{2} + \beta\right) & \dfrac{r}{2} - \beta < |f| < \dfrac{r}{2} + \beta \\ 0 & |f| > \dfrac{r}{2} + \beta \end{cases} \tag{17b}$$

y los filtros de formación de pulsos que se requieren resultan ser relativamente fáciles de sintetizar de manera aproximada. Las correspondientes funciones del tiempo son

$$p(t) = \frac{\cos 2\pi\beta t}{1 - (4\beta t)^2}\,\text{senc}\, rt \tag{18a}$$

$$= \frac{\text{senc}\, 2rt}{1 - (2rt)^2} \qquad \beta = \frac{r}{2} \tag{18b}$$

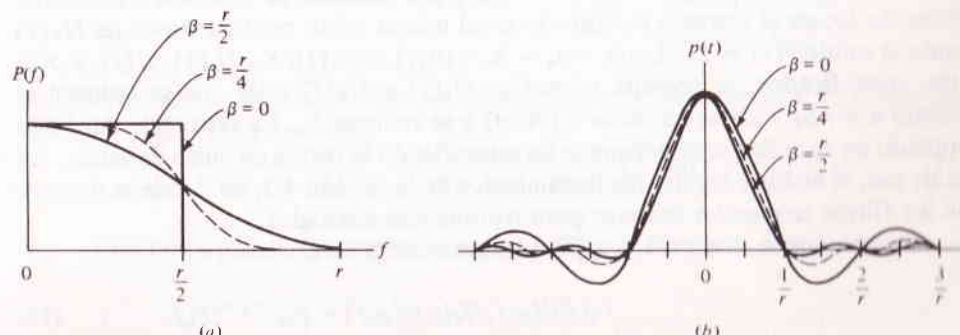



FIGURA 10.4
Formación de pulsos de Nyquist. (*a*) Espectros; (*b*) funciones del tiempo.

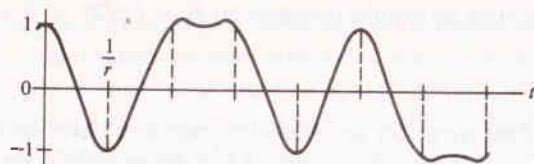



FIGURA 10.5
Forma de onda en banda base para 10110100 empleando pulsos de Nyquist con $\beta = r/2$.

mostradas en la figura 10.4*b* para $\beta = r/4$ y $r/2$ con senc rt. Nótese que las oscilaciones anteriores y posteriores de $V(t)$ decaen en forma más rápida que las de senc rt; esto significa que la sincronización será menos crítica que en un sistema de pulsos seno, puesto que unos modestos errores de temporización no originan grandes cantidades de interferencia intersimbólica. †

Una inspección más a fondo revela otras dos propiedades convenientes de $p(t)$ cuando $\beta = r/2$: el ancho del pulso a la mitad de la amplitud es exactamente igual al espaciamiento entre pulsos $1/r$, y hay cruzamientos cero adicionales en $t = \pm 3/2r$, $\pm 5/2r, \ldots$ En consecuencia, una señal digital polar construida con tales pulsos los centros de pulso dondequiera que haya un cambio de polaridad. La figura 10.5 ilustra esto para el mensaje binario **10110100**. Estos cruzamientos cero se pueden emplear después para generar una ***señal de temporización*** para efectos de sincronización, haciendo obvia la necesidad de transmitir la información de temporización por separado. Nyquist demostró que $p(t)$, por la ecuación (18*b*), es la ***única*** forma de pulso de banda limitada que posee todos estos aspectos convenientes. Pero al precio de una reducción del 50% en la velocidad de señalización, puesto que $r = B$ en vez de $2B$.

Filtros terminales óptimos ★

Ahora considérese el diseño general de un sistema de banda base que emplee pulsos de formación de Nyquist; es decir, que la forma de los pulsos de salida tengan la forma de la ecuación (16). Con referencia a la figura 10.6, la señal de entrada $x(t) = \Sigma_k a_k p_S(t - k/r)$ se filtra por medio de $H_T(f)$ produciendo la potencia transmitida promedio S_T; en el extremo receptor la señal más el ruido pasan a través de $H_R(f)$ dando la salida $y(t) = \Sigma_k A_k p(t - t_d - K/r) n(t)$. si $p_S(t)$, S_T, $H_c(f)$, $G(f)$, y $p(t)$ están especificados, se necesita encontrar $H_T(f)$ y $H_R(f)$ tales que se reduzca al mínimo a $\overline{n^2}/S_T$ y así se aumente a (A/σ) y se reduzca P_e. La estrategia del filtro acoplado no se aplica aquí porque se ha especificado la forma de pulso de salida. En vez de eso, el análisis seguirá los lineamientos de la sección 4.3, en donde se dedujeron los filtros terminales óptimos para transmisión analógica.

Puesto que se dan $p(t)$ y $p_S(t)$, una restricción en el diseño es

$$P_S(f)H_T(f)H_C(f)H_R(f) = K_R e^{-j\omega t_d}P(f) \tag{19}$$

† Por otra parte, $E k \phi m\ A k p[m - k)/r]$ puede ser en realidad *limitada* si $p(t) = \text{senc}(r + \epsilon)t$ en donde ϵ es un error pequeño en la velocidad de señalización.

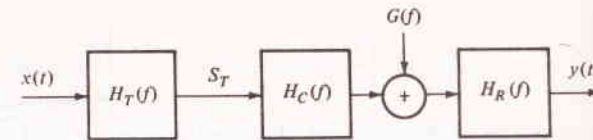



FIGURA 10.6

donde $P(f) = \mathscr{F}[p(t)]$, etc. La segunda restricción comprende la potencia transmitida S_T la cual se calcula encontrando la potencia promedio en una función muestra y tomando después el promedio de ensamble, siendo necesaria esta técnica indirecta porque los pulsos no son necesariamente ortogonales. Prosiguiendo con el cálculo, la señal en la salida de $H_T(f)$ es $\Sigma_k a_k p_T(t - k/r)$ donde $p_T(t) = \mathscr{F}^{-1}[H_T(f)P_s(f)]$, por lo que

$$S_T = \mathbf{E}\left\{\lim_{M\to\infty}\frac{r}{2M}\int_{-M/r}^{M/r}\left[\sum_{k=-M}^{M}a_k\,p_T\left(t-\frac{k}{r}\right)\right]^2 dt\right\}$$

$$= \lim_{M\to\infty}\frac{r}{2M}\sum_{k=-M}^{M}\sum_{m=-M}^{M}\mathbf{E}[a_k a_m]\int_{-M/r}^{M/r}p_T\left(t-\frac{k}{r}\right)p_T\left(t-\frac{m}{r}\right)dt$$

Si los dígitos de fuente son estadísticamente independientes

$$\mathbf{E}[a_k a_m] = \begin{cases}\overline{a^2} & m = k\\ 0 & m \neq k\end{cases}$$

y

$$S_T = r\overline{a^2}\int_{-\infty}^{\infty}p_T^2(t)\,dt = r\overline{a^2}\int_{-\infty}^{\infty}|H_T(f)P_S(f)|^2\,df \tag{20a}$$

en donde, puesto que $A_k = K_R a_k$,

$$\overline{a^2} = \frac{\overline{A^2}}{K_R^2} = \frac{(\mu^2 - 1)A^2}{3K_R^2} \tag{20b}$$

lo cual procede de la ecuación (7) aun cuando S_R puede no ser igual a $\overline{A^2}$. Como para la potencia de ruido de salida,

$$N = \sigma^2 = \overline{n^2} = \int_{-\infty}^{\infty}|H_R(f)|^2 G(f)\,df \tag{21}$$

Combinando las ecuaciones (19), (20) y (21), se elimina a $H_T(f)$ para obtener

$$\left(\frac{A}{\sigma}\right)^2 = \frac{3S_T}{(\mu^2 - 1)r}\left[\int_{-\infty}^{\infty}|H_R(f)|^2 G(f)\,df\int_{-\infty}^{\infty}\frac{|P(f)|^2}{|H_C(f)H_R(f)|^2}\,df\right]^{-1}$$

el producto de las integrales tiene una forma como la de la ecuación (11), (Sec. 4.3).

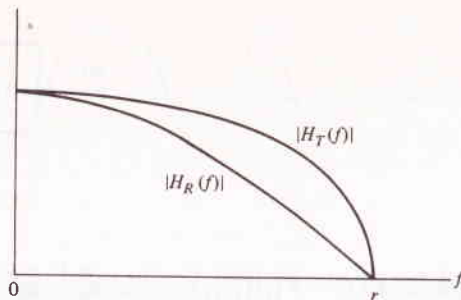



FIGURA 10.7

Empleando la desigualdad de Schwarz en la misma forma, el producto se simplifica cuando

$$|H_R(f)|_{\text{ópt}}^2 = \frac{K|P(f)|}{|H_C(f)|\,G^{1/2}(f)} \tag{22a}$$

$$|H_T(f)|_{\text{ópt}}^2 = \frac{K_R^2|P(f)|\,G^{1/2}(f)}{K|P_S(f)|^2|H_C(f)|} \tag{22b}$$

en donde K es una constante arbitraria. Éstas se pueden comparar con los filtros terminales óptimos para transmisión analógica, ecuación (13), (Sec. 4.3). Como entonces, los corrimientos de fase de los filtros son arbitrarios, siempre y cuando se satisfaga la ecuación (19).

Con los filtros óptimos

$$\left(\frac{A}{\sigma}\right)^2_{\text{máx}} = \frac{3S_T}{(\mu^2-1)r}\left[\int_{-\infty}^{\infty}\frac{|P(f)|\,G^{1/2}(f)}{|H_C(f)|}\,df\right]^{-2} \tag{23}$$

En el caso de ruido blanco y un canal sin distorsión, sustituyendo a $G(f) = \eta/2$ y $|H_C(f)| = K_R$ se tiene

$$\left(\frac{A}{\sigma}\right)^2_{\text{máx}} = \frac{6K_R^2S_T}{(\mu^2-1)\eta r}\left[\int_{-\infty}^{\infty}|P(f)|\,df\right]^{-2} \tag{24}$$

Pero $K_R^2S_T = S_R$ cuando el canal es sin distorsión, y los pulsos de formación de Nyquist, por la ecuación (17), tienen $\int_{-\infty}^{\infty}|P(f)|df = 1$, por lo que

$$\left(\frac{A}{\sigma}\right)^2_{\text{máx}} = \frac{6S_R}{(\mu^2-1)\eta r} = \frac{6}{\mu^2-1}\,p$$

Así, el comportamiento del sistema es de nuevo idéntico a uno ideal de banda base.

EJEMPLO 10.1

Para aclarar estos resultados, tómese el caso de ruido blanco y un canal sin distorción con $p_x(t) = \Pi(t/\tau)$, $\tau = 1/r$, y $p(t)$ dados por la ecuación (18b). Entonces

$$P_S(f) = \frac{1}{r}\text{senc}\left(\frac{f}{r}\right) \qquad P(f) = \frac{1}{r}\cos^2\left(\frac{\pi f}{2r}\right)\Pi\left(\frac{f}{2r}\right)$$

Insertando éstas en la ecuación (22) y prescindiendo de las constantes de proporcionalidad, se tiene

$$|H_R(f)| = \cos\left(\frac{\pi f}{2r}\right)\Pi\left(\frac{f}{2r}\right) \qquad |H_T(f)| = \frac{\cos(\pi f/2r)}{\text{senc}^{1/2}(f/r)}\Pi\left(\frac{f}{2r}\right) \tag{25}$$

como se ve en la gráfica de la figura 10.7. Nótese la ligera elevación de alta frecuencia en $|H_T(f)|$; si $\tau \ll 1/r$, esta elevación es despreciable y $|H_T(f)| \approx |H_R(f)|$ por lo que un diseño sirve para ambos filtros; una ventaja de producción cuando se van a implementar muchos de estos sistemas.

Igualación

Sin importar la forma de pulso que en particular se escoja para la señal digital de banda base, ocurre inevitablemente cierta cantidad de interferencia intersimbólica residual, que obedece a un diseño imperfecto de filtros, a un conocimiento incompleto de las características del canal, etc. En consecuencia, a menudo se inserta un filtro de igualación ajustable entre el filtro de recepción y el convertidor A/D, en particular en sistemas conmutados en donde no se conocen de antemano las características específicas del canal. Tales igualadores que se asemejan a un "trapeador", por lo general toman la forma de una línea de retardo con derivaciones o filtro transversal analizado antes cuando se vio la distorsión lineal en la sección 4.2. Sin embargo, la estrategia para esquematizar este tipo de igualación es algo diferente y merece que se le considere aquí.

La figura 10.8 muestra un igualador transversal con $2M + 1$ derivaciones y un retardo total $2M/r$. Si $p(t)$ es la forma del pulso de entrada, entonces la salida igualada es

$$p_{eq}(t) = \sum_{m=-M}^{M} c_m\, p\left(t - \frac{m+M}{r}\right) \tag{26}$$

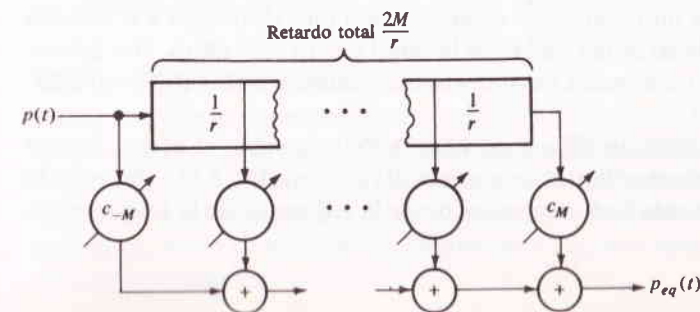



FIGURA 10.8
Igualador transversal.

Suponiendo que $p(t)$ tiene su pico en $t = 0$ e interferencia intersimbólica en ambos lados, a los tiempos de muestreo en la salida se les debe tomar como

$$t_k = \frac{k + M}{r}$$

por lo que

$$p_{eq}(t_k) = \sum_{m=-M}^{M} c_m p\left(\frac{k - m}{r}\right) \tag{27}$$

La ecuación (27) tiene la forma matemática conocida como *convolución discreta*.

En forma ideal, sería bueno tener

$$p_{eq}(t_k) = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases}$$

eliminándose con ello, por completo, la interferencia intersimbólica. Pero esta condición no es posible conseguirla porque sólo se tienen $2M + 1$ variables disponibles, o sean, las ganancias de derivación c_m. Además, la colocación óptima para estas ganancias no es obvia de manera inmediata. Una aproximación es situar los c_m tales que

$$p_{eq}(t_k) = \begin{cases} 1 & k = 0 \\ 0 & k = \pm 1, \pm 2, \ldots, \pm M \end{cases} \tag{28}$$

lo cual, con la ecuación (27), da lugar a $2M + 1$ ecuaciones lineales simultáneas que se pueden resolver par los c_m's. La ecuación (28) describe a un *igualador de forzamiento cero*, puesto que $p_{eq}(t_k)$ tiene M valores cero en cada lado. Esta estrategia es óptima en el sentido de que reduce al mínimo el pico de la interferencia intersimbólica, y tiene la ventaja adicional de la sencillez. Se conocen otras estrategias que son óptimas, pero en un sentido diferente.

Cuando no se conocen de antemano los valores de $p[(k - m)/r]$, las ganancias de derivación se deben situar empleando un proceso iterativo y un criterio de error; es decir, empezando con una posición inicial, se transmite un pulso, se mide el error, se restablecen las derivaciones, y la operación se repite una o más veces. Lo incómodo, de manera obvia, de este procedimiento, lleva aparejado un considerable interés en los igualadores automáticos o de *adaptación* que se ajustan a sí mismos empleando mediciones de error derivadas de la señal efectiva de datos.† La igualación de adaptación tiene un especial valor cuando las características del canal cambian con el tiempo.

La igualación de trapeado (fija o de adaptación) presenta el efecto de que incrementa algo la potencia de ruido en la entrada del convertidor A/D. Pero por lo general, ese efecto queda más que compensado por la reducción de la interferencia intersimbólica.

† En Lucky, Salz y Weldon (1968, Cap. 6) se analizan los algoritmos para ajuste de igualadores y la implementación de la igualación de adaptación.

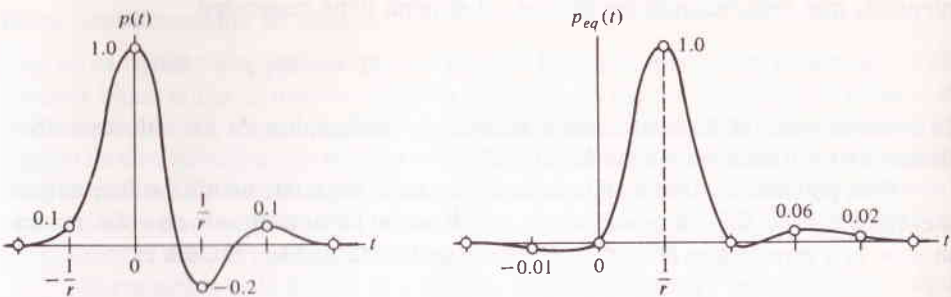



FIGURA 10.9

Ejemplo 10.2

Se va a diseñar un igualador de forzamiento cero y tres derivaciones ($M = 1$) para el pulso distorsionado $p(t)$ de la figura 10.9*a*. Las ganancias de derivación se calculan con facilidad expresando las ecuaciones (27) y (28) en la forma matricial

$$\begin{bmatrix} 1.0 & 0.1 & 0.0 \\ -0.2 & 1.0 & 0.1 \\ 0.1 & -0.2 & 1.0 \end{bmatrix} \begin{bmatrix} c_{-1} \\ c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

Por lo tanto,

$$c_{-1} = -0.096 \qquad c_0 = 0.96 \qquad c_1 = 0.2$$

y los correspondientes valores muestra $p_{eq}(t)$ se pueden ver en la gráfica, en la figura 10.9*b* con una curva interpolada. Como se esperaba, existe un cero a cada lado del pico. Sin embargo, el forzamiento cero ha producido alguna pequeña interferencia intersimbólica en puntos muy afuera de donde el pulso no igualado fue cero.

Señalización binaria contra μ-aria

Hasta ahora se han tomado los valores de r y μ conforme los dicta la fuente. Pero esta limitación no debe ser motivo de preocupación mientras que la velocidad de información en la transmisión surta los datos fuente. Un conocimiento superficial de la teoría de la información ayuda a poner esto en forma cuantitativa. De manera específica, si los parámetros de fuente son r y μ_s, entonces la velocidad de información de la fuente es

$$\mathcal{R}_S = r_S \log_2 \mu_S \qquad \text{bits/s}$$

mientras, que despreciando los errores, el sistema tiene capacidad

$$\mathscr{C} = r \log_2 \mu \qquad \text{bits/s}$$

En consecuencia, el diseñador debe seleccionar cualesquier de los valores convenientes de r y μ siempre y cuando $\mathscr{C} \geq \mathscr{R}_S$.

Si la potencia se tiene en alta estima y no así el ancho de banda, se debe emplear el valor *más pequeño* de μ —es decir, señalización *binaria*. Puesto que $\mathscr{C} = r$ cuando $\mu = 2$, y puesto que $B \geq r/2$, el ancho de banda que se requiere es

$$B \geq \tfrac{1}{2}\mathscr{R}_S \tag{29}$$

La extensión en que se excede el ancho de banda disponible de $\mathscr{R}_s/2$, indica la cantidad de flexibilidad de que dispone el diseñador en la selección de la forma de pulso. Suponiendo ruido blanco gaussiano y filtros óptimos para la elección de la forma de pulso, el requisito de potencia de la señal se calcula de la probabilidad deseada de error P_e por medio de

$$Q\left(\sqrt{\frac{2S_R}{\eta\mathscr{R}_S}}\right) \leq P_e \tag{30}$$

lo cual procede de las ecuaciones (8) a la (10) con $\mu = 2$ y $r = \mathscr{R}_S$.

Por otra parte, el ancho de banda se reduce al mínimo si se toma a μ como el valor *más grande* posible, puesto que $r = \mathscr{R}_s/\log_2\mu$. En conformidad, se debe encontrar el mayor valor de μ tal que

$$2\left(1 - \frac{1}{\mu}\right)Q\left(\sqrt{\frac{6 \log_2 \mu}{\mu^2 - 1}\frac{S_R}{\eta\mathscr{R}_S}}\right) \leq P_e \tag{31}$$

Entonces

$$B \geq \frac{\mathscr{R}_S}{2 \log_2 \mu_{\text{máx}}} \tag{32}$$

en donde para la obtención del límite inferior se incluyen pulsos senc.

En cualquier caso, se debe dar una traslación apropiada de código en transmisor y receptor siempre que $\mu \neq \mu_s$. Esta codificación sólo sirve para crear un mejor acoplamiento entre la fuente y el canal, y no se debe confundir con la codificación para control de error o la codificación que disminuye la redundancia en los mensajes de fuente.

EJERCICIO 10.2 Un computador determinado produce dígitos octales ($\mu_s = 8$) a una velocidad de 10 000 por segundo. Si el ancho de banda disponible es de 20 kHz y $\eta = 5 \times 10^{-8}$ W/Hz, selecciónense los parámetros apropiados del sistema tales que $P_e \leq 10^{-4}$. *Respuesta*: $\mu = 2$ y $r = 30\,000$; la forma de pulso la da la ecuación (18*a*) con $\beta \leq 2B - r = 10$ kHz; $S_R \geq (3.7)^2\, \eta\, \mathscr{R}_S/2 \sim 2.1$ mW.

Otras consideraciones de diseño

Hay otros numerosos factores que, de manera bastante apropiada, deben ser considerados importantes al diseñar sistemas prácticos de banda base. Uno de éstos es el *ruido de impulso*, o sean, los pulsos esporádicos causados por tormentas eléctricas y equipo de conmutación que pueden constituir un problema particularmente molesto para la transmisión de datos por medio de circuitos telefónicos. Otros factores incluyen métodos de sincronización, remoción y restauración de CD para sistemas de acoplamiento de CA, y diferentes técnicas de codificación de línea.

Sin embargo, cada uno es en sí mismo, un tema especial y se deben omitir aquí en atención a los de interés más general. El estudiante encontrará suficiente material de literatura profesional en la lista de lectura suplementaria seleccionada, la cual constituye una guía y punto de partida.

10.2 MODULACIÓN DIGITAL: (ASK), (FSK) Y (PSK)

Así como existe una multitud de técnicas de modulación para señales analógicas, también la información digital se puede imprimir sobre una onda portadora de

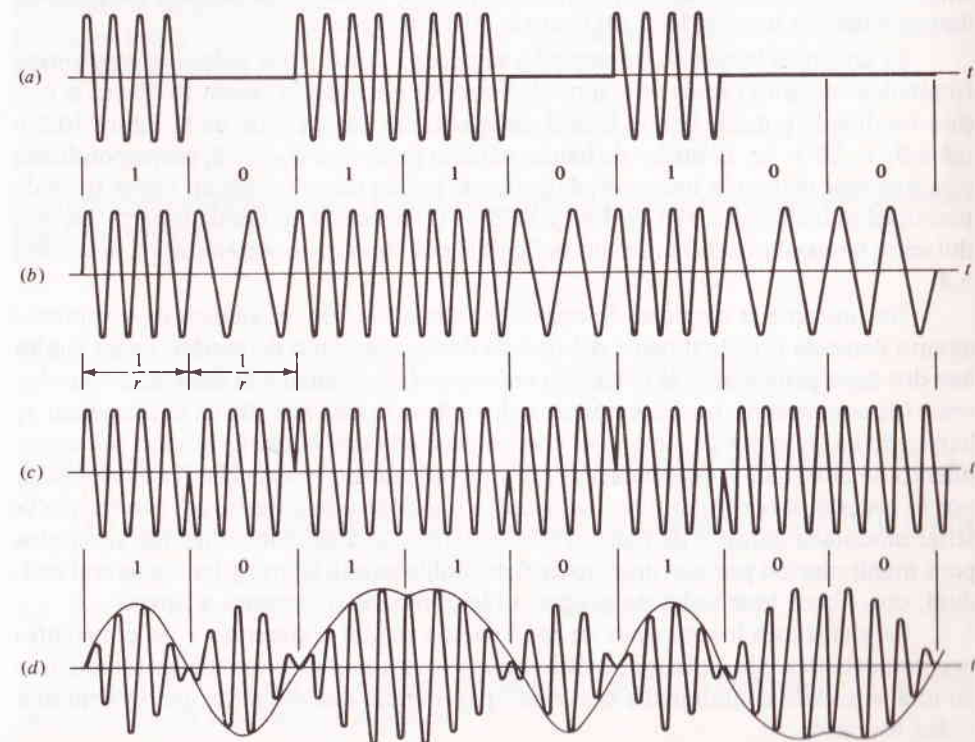



FIGURA 10.10
Formas de onda de modulación digital para el mensaje binario 10110100 (*a*) Manipulación por corrimiento de amplitud; (*b*) manipulación por corrimiento de frecuencia; (*c*) manipulación por corrimiento de fase; (*d*) doble banda lateral con formación de pulsos en banda base.

muchas maneras. Esta sección se ocupa de los tipos básicos y algunas de sus variantes, recurriendo a los resultados de la sección anterior aunados a la teoría de la información de los capítulos 5 al 7. Se recalcará principalmente lo relativo al comportamiento del sistema en presencia del ruido, es decir, las probabilidades de error como una función de ρ. En la mayor parte se concentrará la atención en las señales binarias; la extensión a las señales μ-arias no es en forma conceptual difícil, pero ello involucra el empleo de matemáticas más arduas.

Dado un mensaje digital, la técnica de modulación más sencilla es la *manipulación por corrimiento de amplitud* (ASK), donde la amplitud de la portadora se conmuta entre dos o más valores, por lo general el *puesto* (on) y el *fuera* (off) de las señales binarias. La onda modulada resultante consiste entonces de pulsos de RF o *marcas*, que representan al binario **1**, y *espacios*, que representan al binario **0** (Fig. 10.10*a*). De manera similar, se podría manipular la frecuencia o la fase (Figs. 10.10*b* y *c*) dando la *manipulación por corrimiento de frecuencias* (FSK) o la *manipulación por corrimiento de fase* (PSK). Estos tipos de modulación corresponden a la AM, FM y PM respectivamente, con una señal modulante de pulso rectangular. Claro que el precio de esta sencillez es una ancho de banda de transmisión excesivo; es decir, $B_T \gg r$. La manipulación por corrimiento de amplitud también tiene potencia desperdiciada en la portadora, tal como la AM analógica.

El ancho de banda de transmisión se puede reducir si los pulsos se *conforman* (limitados en banda) antes de la modulación. Por ejemplo, la figura 10.10*d* es la versión modulada a doble banda lateral de la señal de banda base de la figura 10.5 y tiene $B_T = 2B = 2r$. El ancho de banda mínimo posible es $B_T \approx r/2$, correspondiente a pulsos senc en banda base y modulación de banda lateral residual. Otros tipos de modulación incluyen la FM, la PM y VSB + C formados en banda base. † La modulación de banda lateral única no es factible por las razones expuestas en la sección 5.4.

No obstante la cantidad de opciones de modulación, el análisis de comportamiento depende principalmente del tipo de demodulación o detección, de las cuales hay dos tipos principales: la detección *coherente* (sincrónica) y la detección *no coherente* (de envolvente). La detección no coherente es la más sencilla de implementar y, hablando en términos generales, se emplea asociada con la manipulación por corrimiento de amplitud y manipulación por corrimiento de frecuencia. Tales sistemas, por lo general, incorporan filtros acoplados (de integrado y descarga) puesto que la señal modulada consiste de pulsos rectangulares. La detección coherente se emplea para manipulación por corrimiento de fase, doble banda lateral y banda lateral residual, con filtros terminales acoplados o bien óptimos o cercanos a éstos.

Se analizarán los sistemas de modulación digital coherentes y no coherentes más importantes. Sin embargo, es útil primero examinar las propiedades estadísticas de una sinusoide contaminada con ruido pasabanda, esa situación que es común a todos los casos.

† Croisler y Pierret (1970) han ideado una alternativa prometedora que ellos designan como *modulación digital de eco.*

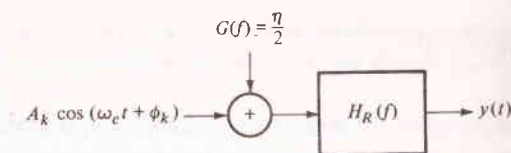



FIGURA 10.11

Envolvente y fase de una sinusoide más ruido pasabanda

Considérese un sistema de modulación digital con frecuencia portadora f_c. En el receptor, la señal se contamina con ruido blanco, después se pasa a través de un filtro pasabanda $H_R(f)$ centrado en f_c, figura 10.11. El ruido pasabanda de salida es, de la sección 7.3

$$n(t) = n_i(t) \cos(\omega_c t + \theta) - n_q(t) \operatorname{sen}(\omega_c t + \theta) \qquad (1a)$$

$$= R_n(t) \cos[\omega_c t + \theta + \phi_n(t)] \qquad (1b)$$

en donde se ha incluido una fase constante arbitraria θ. Recuérdese que las componentes en fase y cuadratura $n_i(t)$ y $n_q(t)$ son variadas gaussianas de valor medio cero independientes con

$$\overline{n_i^2} = \overline{n_q^2} = \overline{n^2} = N = \frac{\eta}{2}\int_{-\infty}^{\infty} |H_R(f)|^2\, df \qquad (2)$$

mientras que la envolvente $R_n = \sqrt{n_i^2 + n_q^2}$ tiene la función de densidad de probabilidad de *Rayleigh*

$$P_{R_n}(R_n) = \frac{R_n}{N} e^{-R_n^2/2N} \qquad R_n \geq 0$$

$$\overline{R_n} = \sqrt{\frac{\pi N}{2}} \qquad \overline{R_n^2} = 2N \qquad (3)$$

en tanto que la fase $\phi_n = \arctan(n_i/n_q)$ está uniformemente distribuida sobre $|\phi_n| \leq \pi$.

En ausencia de ruido, sea la señal modulada en la salida de $H_R(f)$ de la forma

$$K_R x_c(t) = A_k \cos(\omega_c t + \phi_k) \qquad t = \frac{k}{r}$$

donde A_k y/o ϕ_k representan el dígito de mensaje k-ésimo. Tomando a $\theta = \phi_k$ en la ecuación (1), la señal más el ruido es

$$y(t) = [A_k + n_i(t)] \cos(\omega_c t + \phi_k) - n_q(t) \operatorname{sen}(\omega_c t + \phi_k) \qquad (4a)$$

$$= R(t) \cos[\omega_c t + \phi_k + \phi(t)] \qquad (4b)$$

en donde

$$R^2 = (A_k + n_i)^2 + n_q^2 \qquad \phi = \arctan \frac{n_q}{A_k + n_i} \tag{5}$$

La ecuación (5) relaciona la envolvente y la fase de una sinusoide más ruido pasabanda con la señal y los componentes de ruido, y se tiene interés de manera específica en las funciones de densidad de probabilidad de R y ϕ.

Antes de profundizar en detalles, especúlese sobre la naturaleza de $p_R(R)$ y $p_\phi(\phi)$ bajo condiciones extremas. Cuando $A_k = 0$, R y ϕ se reducen a R_n y ϕ_n, la envolvente y la fase del ruido solo, los cuales tienen funciones de densidad de probabilidad de Rayleigh y uniforme, respectivamente. En el otro extremo, si $A_k \gg \sqrt{N}$, $R \approx A_k + n_i$ y $\phi \approx n_q/A_k$ por lo que ambas serán aproximadamente gaussianas. Puesto que A_k^2 es proporcional a la potencia de la señal S_R, este caso corresponde a $S_R/N \gg 1$.

Para casos intermedios se sigue el procedimiento que condujo a la ecuación (11), (Sec. 7.3), sustituyendo a n_i por $A_k + \nu_i$. Esto origina la función de densidad de probabilidad conjunta

$$p(R,\phi) = \frac{R}{2\pi N} e^{-(R^2 - 2A_kR\cos\phi + A_k^2)/2N} \tag{6}$$

con $R \geq 0$ y $|\phi| \leq \pi$ por definición. Diferente al caso del ruido pasabanda solo, el término producto $R \cos \phi$ en la ecuación (6) significa que R y ϕ no son estadísticamente independientes. La separación de $p_R(R)$ y $p_\phi(\phi)$ requiere, por lo tanto, integración, es decir,

$$\begin{aligned} p_R(R) &= \int_{-\pi}^{\pi} p(R,\phi)\, d\phi \\ &= \frac{R}{N} e^{-(R^2 + A_k^2)/2N} \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{(A_kR/N)\cos\phi}\, d\phi \\ &= \frac{R}{N} e^{-(R^2 + A_k^2)/2N} I_0\left(\frac{A_k R}{N}\right) \qquad R \geq 0 \end{aligned} \tag{7}$$

en donde I_0 es la función de Bessel modificada de primera clase y de orden cero. La ecuación (7) se conoce como la *distribución de Rice*; su apariencia formidable quizá sea desalentadora, y además no es una función trivial. Por fortuna, muchos sistemas satisfarán la condición de señal grande $A_k \gg \sqrt{N}$ para la cual $I_0(A_kR/N) \approx \sqrt{N/2\pi A_k R}\, e^{A_kR/N}$. Por lo que, si $S_R/N \gg 1$,

$$p_R(R) \approx \sqrt{\frac{R}{2\pi A_k N}}\, e^{-(R - A_k)^2/2N} \qquad R \geq 0 \tag{8}$$

la cual es en forma esencial gaussiana con $\overline{R} = A_k$ y $\sigma_R^2 = N$. Esa condición procede del hecho de que $R/2\pi A_k N \approx 1/2\pi N$ en la vecindad de $R = A_k$ en donde $p_R(R)$ tiene la mayor parte de su área.

Volviendo a la fase, se puede encontrar a $p_\phi(\phi)$ de manera exacta, integrando la ecuación (6) sobre $R = 0$ *a* ∞; la integración es algo larga y el resultado es analíticamente engorroso —(véase el Prob. 10.14). Consistente con $S_R/N \gg 1$, ello simplifica a

$$p_\phi(\phi) \approx \sqrt{\frac{A_k^2}{2\pi N}} \cos\phi\, e^{-(A_k \operatorname{sen}\phi)^2/2N} \qquad |\phi| \leq \frac{\pi}{2} \tag{9}$$

la cual para valores pequeños de ϕ, se aproxima a gaussiana con $\overline{\phi} = 0$ y $\overline{\phi^2} = N/A_k^2$. La ecuación (9) se invalida para $|\phi| > \pi/2$ (¿por qué?), pero la probabilidad de ese evento es pequeña pues tiende a desaparecer bajo la condición supuesta.

Manipulación por corrimiento de amplitud no coherente

Poniendo los resultados anteriores a trabajar, se empieza con la manipulación por corrimiento de amplitud binaria no coherente. Se supone que las formas de onda recibidas se asemejan a la figura 10.10*a* con $\tau = 1/r$ y $f_c \gg r$. Si E_R es la energía en cada pulso marca recibido (representando a 1), entonces la potencia promedio de la señal es $S_R = 1/2rE_R$ puesto que las marcas y los espacios son igualmente probables.

El sistema de detección de un filtro pasabanda (o uno de integrado y descarga) acoplado a los pulsos marca de RF, es decir,

$$h_R(t) = \cos \omega_c t\, \Pi[r(t - t_0)] \tag{10}$$

seguido de un detector de envolvente y un convertidor A/D, (Fig. 10.12). Cuando se recibe un pulso marca, la salida filtrada es un pulso triangular de RF con valor pico

$$A = \sqrt{\frac{E_R}{2r}} = \sqrt{\frac{S_R}{r^2}} \tag{11}$$

Así, en los tiempos óptimos de muestreo, la envolvente tiene $A_k = A$ o 0 (el no pulso o la salida espacio). La potencia de ruido filtrada es

$$N = \frac{\eta}{2} \int_{-\infty}^{\infty} |H_R(f)|^2\, df = \frac{\eta}{2} \int_{-\infty}^{\infty} h_R^2(t)\, dt = \frac{\eta}{4r} \tag{12}$$

por lo que

$$\left(\frac{A^2}{N}\right)_{\text{máx}} = \frac{2E_R}{\eta} = \frac{4S_R}{\eta r} = 4\rho \tag{13}$$

en donde $\rho = S_R/\eta r$ como antes.

El cálculo de la probabilidad de error comprende dos funciones de densidad de probabilidad: cuando se envía un **0** o un espacio, la envolvente resultante tiene una función de densidad de probabilidad de Rayleigh, $p(R/A_k = 0) = pR_n(R_n)$; cuan-

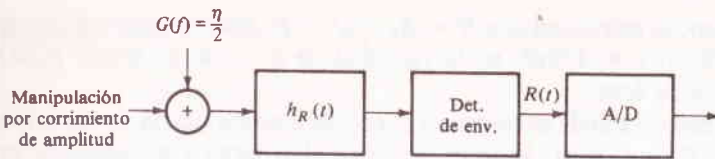



FIGURA 10.12
Detección no coherente de manipulación por corrimiento de amplitud.

do se envía un **1** o una marca, $p(R/A_k = A)$ tiene una función de densidad de probabilidad de Rice. La figura 10.13 muestra las dos funciones de densidad, suponiendo $\rho \gg 1$ por lo que se puede emplear la ecuación (8) para $p(R/A_k = A)$. Es evidente que el nivel de umbral de decisión del convertidor A/D se debe situar entre 0 y A, pero el mejor valor no es necesariamente $A/2$. En relación con esto, no existe un umbral "mejor" en el sentido de igualar a P_{e_1} y a P_{e_0} y reducir al mínimo en forma simultánea a P_e. El mínimo de P_e se consigue tomando el umbral donde se intersecan las dos curvas (¡explíquese esta aseveración!), lo cual resulta ser en forma aproximada $(A/2)\sqrt{1 + (2/\rho)}$.

Por lo general, el umbral se sitúa en $A/2$ lo cual es cercano a lo óptimo si $\rho \gg 1$. Entonces

$$P_{e_0} = \int_{A/2}^{\infty} p_{R_n}(R_n)\, dR_n = e^{-A^2/8N} = e^{-\rho/2} \tag{14a}$$

mientras que

$$P_{e_1} = \int_0^{A/2} p_R(R)\, dR \approx Q\left(\frac{A}{2\sqrt{N}}\right) = Q(\sqrt{\rho})$$

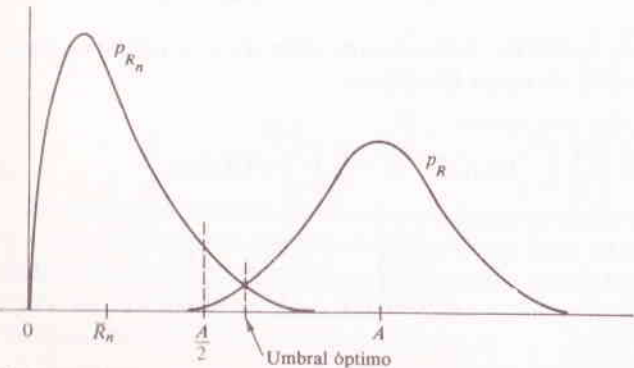



FIGURA 10.13
Funciones de densidad de probabilidad para detección no coherente de la manipulación por corrimiento de amplitud.

en donde la aproximación es el área de $-\infty$ a $A/2$ de una gaussiana con $\overline{R} = A$ y $\sigma_R^2 = N$. Introduciendo la expresión asintótica para $Q(\sqrt{\rho})$, ecuación (10), (Sec. 3.5), da

$$P_{e_1} \approx \frac{1}{\sqrt{2\pi\rho}}\, e^{-\rho/2} \tag{14b}$$

resultando así el hecho de que $P_{e_1} \ll P_{e_0}$ cuando $\rho \gg 1$. Por último,

$$P_e = P_0 P_{e_0} + P_1 P_{e_1}$$

$$\approx \frac{1}{2}\left(1 + \frac{1}{\sqrt{2\pi\rho}}\right) e^{-\rho/2} \tag{15}$$

Para comparar este resultado con la transmisión de banda base con las mismas condiciones, se tiene de la ecuación (8), (Sec. 10.1), con $\rho \gg 1$,

$$P_{e_{BB}} = Q(\sqrt{2\rho}) \approx \frac{1}{2\sqrt{\pi\rho}}\, e^{-\rho} \tag{16}$$

Por lo tanto,

$$P_{e_{ASK}} \approx \sqrt{\pi\rho}\, e^{+\rho/2}\, P_{e_{BB}}$$

así, si $\rho = 10$, la probabilidad de error en la manipulación por corrimiento de amplitud es alrededor de 800 veces la de un sistema de banda base bien diseñado y casi todos los errores radican en cambios de **0**s a **1**s. Es obvio que la manipulación por corrimiento de amplitud en comparación no es muy favorable. Además, como en la mayor parte de lo estudiado con anterioridad, la ecuación (15) representa, de manera virtual, lo mejor que se puede lograr en este sistema. Un diseño de filtros imperfectos, una sincronización pobre, etc., todo contribuye al incremento de P_e. En particular, los filtros acoplados pasabanda son difíciles de construir, por lo que los sistemas prácticos por lo general

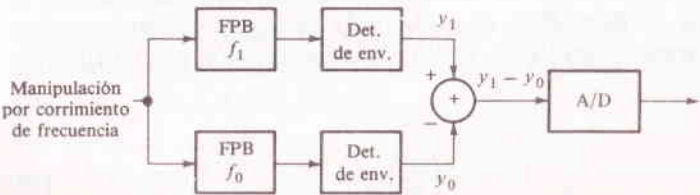



FIGURA 10.14
Detección no coherente de manipulación por corrimiento de frecuencia binaria.

enen un filtro *de banda base* de integrado y descarga después del detector de envol-
ente, que degrada aún más el comportamiento. Pero en tal caso, no se debe esperar
ucho de un sistema que tiene componentes que no son relativamente complejos.

JERCICIO 10.3 Desarróllese el cálculo de la ecuación (12) y demuéstrese que la
1) es correcta, sustituyendo el valor de N en $A^2/N = 2E_2/\eta$, lo cual siempre es vá-
lo para un filtro acoplado.

lanipulación por corrimiento de frecuencia no coherente

uede parecer extraño que la detección de envolvente funcione en la manipulación
or corrimiento de frecuencia tan bien como en la manipulación por corrimiento de
nplitud. Sin embargo, un cuidadoso examen de la señal binaria de manipulación
or corrimiento de frecuencia de la figura 10.10*b* revela que se debe, en forma bási-
, a dos señales de manipulación por corrimiento de amplitud interpuestas de fre-
uencias portadoras diferentes, digamos f_1 y f_0. De conformidad, se puede realizar la
etección no coherente empleando un par de filtros acoplados y detectores de envol-
ente arreglados de acuerdo a la figura 10.14, con una rama que responde a los pul-
s a f_1, la otra a f_0. Para evitar la diafonía en los tiempos de muestreo, es necesario
ue

$$|f_1 - f_0| = Mr$$

onde M es un entero, que por lo general, se toma como la unidad para reducir al
ínimo al ancho de banda. (Desafortunadamente de esta condición resulta una se-
al que tiene componentes sinusoidales de frecuencia discreta, que pueden tener
fecto contrario en ciertas aplicaciones.)

En ausencia de ruido y diafonía, $y_1 = A$ y $y_0 = 0$ cuando la frecuencia recibida
f_1, y viceversa para f_0. Puesto que la diferencia de envolvente $y_1 - y_0$ es la entrada
el convertidor A/D, se debe situar al nivel de umbral en cero, sin importar el valor
A. Así, $P_{e_1} = P(y_1 - y_0 < 0)$ y, por simetría, $P_{e_0} = P_{e_1} = P_e$, por lo que

$$P_e = P(y_0 > y_1) \tag{17}$$

onde, en base a lo estudiado acerca de la manipulación por corrimiento de ampli-
d, y_0 tiene una función de densidad de probabilidad de Rayleigh mientras que y_1
ene su correspondiente de Rice. Pero como la onda de manipulación por corri-
iento de frecuencia carece de los "espacios" de la manipulación por corrimiento
amplitud, $S_R = rE_R$.

$$\left(\frac{A^2}{N}\right)_{\text{máx}} = \frac{2E_R}{\eta} = \frac{2S_R}{\eta r} = 2\rho \tag{18}$$

la salida de uno u otro filtro, dependiendo de la frecuencia de entrada.

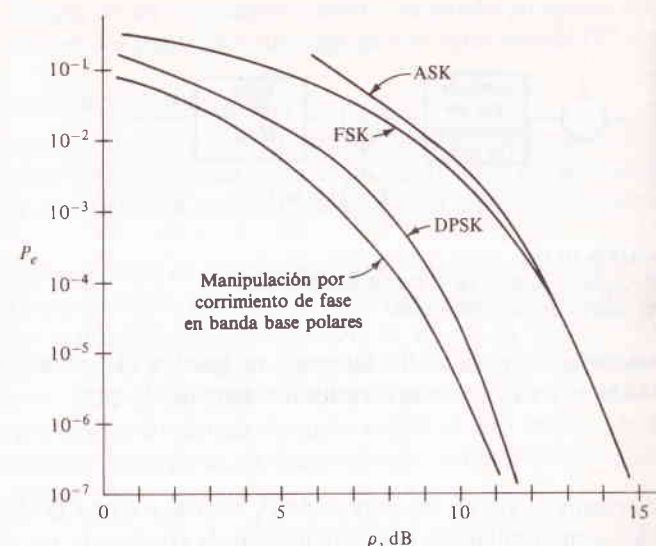



FIGURA 10.15
Probabilidades de error para sistemas binarios de modulación digital.

Para calcular la probabilidad de la ecuación (17), primero obsérvese que $P_{y_0}(y_0) = P_{R_n}(y_0)$ por lo que, para un valor dado de y_1,

$$P(y_0 > y_1 \mid y_1) = \int_{y_1}^{\infty} p_{R_n}(y_0)\, dy_0 = e^{-y_1^2/2N}$$

Entonces, calculando para todos los valores posibles de y_1,

$$P_e = \int_0^{\infty} P(y_0 > y_1 \mid y_1) p_R(y_1)\, dy_1$$

$$= \int_0^{\infty} e^{-y_1^2/2N} \frac{y_1}{N} e^{-(y_1^2 + A^2)/2N} I_0\left(\frac{Ay_1}{N}\right) dy_1$$

en la cual, por la ecuación (7) $p_{y_1}(y_1) = p_r(y_1)$ con $A_K = A$. Esta integral, bastante extraña, puede evaluarse en forma exacta, sin aproximaciones. Para este objeto, sean $\lambda = \sqrt{2}y_1$ y $A/\sqrt{2}$, dando

$$P_e = \frac{1}{2} e^{-A^2/4N} \int_0^{\infty} \frac{\lambda}{N} e^{-(\lambda^2 + a^2)/2N} I_0\left(\frac{a\lambda}{N}\right) d\lambda$$

y la comparación con la ecuación (7) revela que el integrado tiene ahora de manera exacta la misma forma que una función de densidad de probabilidad de Rice; en

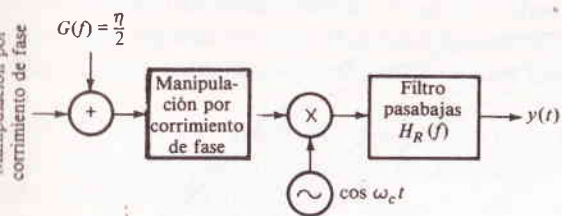



FIGURA 10.16
Detección sincrónica de la manipulación por corrimiento de fase sincrónica.

consecuencia, el área (la integral) es igual a la unidad. Por último, puesto que $A^2/4N = \rho/P$, se tiene el resultado sencillo de que

$$P_e = \tfrac{1}{2}e^{-\rho/2} \qquad (19)$$

Así, como se vio en las curvas de P_e contra ρ cuyas gráficas aparecen en la figura 10.5, la manipulación por corrimiento de frecuencia no coherente no proporciona mejores velocidades de error que la manipulación por corrimiento de amplitud, excepto a valores relativamente pequeños de ρ.

¿Cuáles son entonces las ventajas de la manipulación por corrimiento de frecuencia, si existen algunas? A modo de reflexión adicional, se hacen evidentes tres ventajas sobre la manipulación por corrimiento de amplitud:

1 La manipulación por corrimiento de frecuencia tiene la propiedad de amplitud constante cuyos méritos se trataron con la modulación exponencial analógica.

2 Las probabilidades de error por dígito P_{e1} y P_{e0} son iguales.

3 El nivel de umbral óptimo es independiente de A y ρ, y no se necesitan reajustar si la intensidad de la señal varía con el tiempo, un fenómeno común en la transmisión por radio.

Es precisamente por esta última razón que la manipulación por corrimiento de frecuencia se prefiere a la manipulación por corrimiento de amplitud en aplicaciones donde se espera que haya desvanecimiento y no es factible la detección sincrónica.

Ahora falta de analizarse un último punto: por qué la manipulación por corrimiento de frecuencia binaria no da la reducción de ruido de banda ancha asociada por lo general con la FM. Para el sistema descrito aquí, la razón es muy sencilla: éste es de manera básica más bien AM que FM. Si se emplea en verdad FM, incluyendo detección por medio de un limitador-discriminador,† resulta que el ancho de banda de transmisión que se requiere para controlar la interferencia intersimbólica es tan grande que el efecto de reducción de ruido se cancela esencialmente por el ruido incrementado de predetección. Sin embargo, para señales multinivel ($\mu > 2$) es posible alguna ventaja en la forma de un intercambio ancho de banda-potencia; después de todo, si $\mu \to \infty$, las señales digitales y analógicas son equivalentes.

† Véase Bennett y Davey (1965, Cap. 9).

Manipulación por corrimiento de fase coherente

La detección coherente o sincrónica se apoya en un conocimiento preciso de la fase de la onda portadora recibida, así como de su frecuencia, implicando así el empleo de componentes físicos más complejos. A cambio, esa detección ofrece una mejoría en el comportamiento.

Considérese, por ejemplo, la señal de manipulación por corrimiento de fase binaria de la figura 10.10*c*, o para el objeto, la señal de doble banda lateral de la figura 10.10*d*, puesto que de ambas se puede decir que tienen $A_K = \pm A$ o $\phi_K = 0$ o π. En la figura 10.16 se muestra un detector sincrónico apropiado, donde el oscilador local está sincronizado y $H_r(f)$ es un filtro *pasabajas*; el filtro pasabanda que está antes del mezclador sólo sirve para prevenir sobrecargas de ruido excesivo. Del estudio de la detección sincrónica para señales analógicas, se desprende en forma inmediata que la salida filtrada es de la forma

$$y(t) = \pm A + n_i(t) \qquad (20a)$$

y, con filtraje óptimo,

$$\left(\frac{A^2}{\overline{n_i^2}}\right)_{\text{máx}} = \frac{2S_R}{\eta r} = 2\rho \qquad (20b)$$

Por lo tanto, puesto que $n_i(t)$ es gaussiana,

$$P_e = Q\left(\frac{A}{\sqrt{\overline{n_i^2}}}\right) = Q(\sqrt{2\rho}) \qquad (21)$$

lo mismo que en transmisión de banda base óptima y, por la figura 10.15, 3 a 4 dB mejor que en la modulación no coherente.

Dada la calidad de su comportamiento, combinada con la propiedad de amplitud constante, la manipulación por corrimiento de fase *multinivel* ($\mu > 2$) tiene un valor práctico considerable. Nótese de manera especial, que incrementado a μ se incrementa la velocidad de bits equivalente, sin necesidad de un ancho de banda de transmisión mayor. Es conveniente así gastar un poco de tiempo en la manipulación

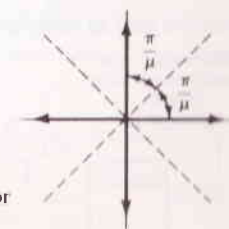



FIGURA 10.17
Fasores y umbrales de la manipulación por corrimiento de fase con $\mu = 4$.

corrimiento de fase μ-aria en donde

$$K_R x_c(t) = A \cos(\omega_c t + \phi_k) \qquad \phi_k = 0, \frac{2\pi}{\mu}, \frac{4\pi}{\mu}, \ldots, \frac{2(\mu - 1)\pi}{\mu} \tag{22}$$

manera conceptual, tal señal se detecta por medio de discriminación de fase con ulos de decisión π/μ, $3\pi/\mu$, . . . , centrados entre los valores esperados de fase de ladora. La figura 10.17 muestra un diagrama fasorial para manipulación por imiento de fase cuaternaria ($\mu = 4$) y los correspondientes umbrales angulares.

La señal recibida más ruido, después del filtraje pasabanda, tiene la forma de cuación (4b), por lo que los errores ocurren siempre que la perturbación de fase icida de ruido $\phi(t)$ cruza el umbral. Por tanto, por simetría, las probabilidades rror por dígitos son iguales y

$$P_e = P\left(|\phi| > \frac{\pi}{\mu}\right) = 1 - \int_{-\pi/\mu}^{\pi/\mu} p_\phi(\phi)\, d\phi \tag{23a}$$

oniendo filtraje pasabanda óptimo, $A^2/N = 2\rho$ por lo que si $\rho \gg 1$,

$$\int_{-\pi/\mu}^{\pi/\mu} p_\phi(\phi)\, d\phi = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{2\rho}\,\operatorname{sen}\pi/\mu} e^{-\lambda^2/2}\, d\lambda \tag{23b}$$

onde se ha empleado la ecuación (9) y hecho a $\lambda = \sqrt{2\rho}$ sen ϕ. Consistente con ondición de señal grande se obtiene por último

$$P_e \approx \frac{1}{\sqrt{\pi\rho}\,\operatorname{sen}\pi/\mu}\, e^{-\rho\,\operatorname{sen}^2 \pi/\mu} \tag{24}$$

ual es válido para $\mu \geq 4$ y $\rho \gg 1$.

ipulación por corrimiento de fase coherente diferencial

a ideado una técnica apropiada conocida como *comparación de fase* o *manipu-n por corrimiento de fase coherente diferencial* (DPSK) para obviar los proble-de sincronización de la detección coherente. La estrategia se muestra en el rama de la figura 10.18 para la manipulación por corrimiento de fase coherente encial binaria; en forma parecida a la detección homodina, el oscilador local se tuye por la propia señal retardada en tiempo en exactamente $1/r$, o sea el espa-

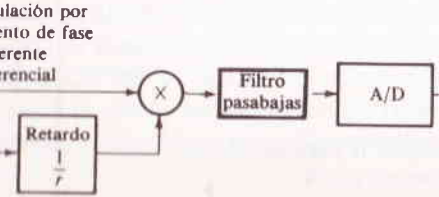



FIGURA 10.18
Detección por comparación de fase de la manipulación por corrimiento de fase binaria coherente diferencial.

ciamiento entre bits. Si los dígitos adyacentes son de fase semejante, su producto da por resultado una salida positiva (o un 1 binario); en forma inversa, las fases opuestas dan por resultado una salida negativa (un **0** binario). Así, es el *corrimiento* o el *no corrimiento* entre los valores de fase transmitidos, lo que representa la información del mensaje, por lo que se requiere en el transmisor, una apropiada codificación, a la que se le designa como *codificación diferencial.*

La codificación diferencial empieza con un primer dígito arbitrario; en seguida indica los dígitos de mensaje por medio de transición sucesiva o no transición. Se establece una transición para el mensaje **0**, y una no transición para el mensaje **1**. El proceso de codificación se muestra a continuación.

Mensaje de entrada		1	0	1	1	0	1	0	0
Mensaje codificado	1	1	0	0	0	1	1	0	1
Fase transmitida	0	0	π	π	π	0	0	π	0
Salida de comparación de fase		+	−	+	+	−	+	−	−
Mensaje de salida		1	0	1	1	0	1	0	0

La codificación diferencial se emplea más a menudo para sistemas de manipulación por corrimiento de fase, pero no se le restringe a tales aplicaciones. En general, es ventajosa en sistemas que no tienen sentido de polaridad absoluta.

En cuanto a la producción de ruido en la manipulación por corrimiento de fase coherente diferencial, podría aparecer que la detección diferencial requiere dos veces más potencia que la detección coherente, dado que la propia referencia de fase se contamina con ruido. Sin embargo, las perturbaciones en realidad tienden a cancelarse en el proceso de comparación, por lo que la degradación no es muy grande. Se desarrollará el análisis para el caso en que las dos fases adyacentes son iguales, es decir, $\phi_K = \phi_{K-1} = 0$, con lo que ocurre un error si la salida es negativa. Escribiendo la señal recibida (y filtrada) más ruido como $[A + n_i(t)] \cos \omega_c t - n_q(t) \operatorname{sen} \omega_c t$, la salida de la unidad de retardo será $[A + n_i(t')] \cos \omega_c t' - n_q(t') \operatorname{sen} \omega_c t'$ donde $t' = t - 1/r$. Así, siempre y cuando f_c sea un entero múltiplo de r (lo cual se puede garantizar por heterodinación si es necesario), la entrada del convertidor A/D será proporcional a

$$y(t) = [A + n_i(t)][A + n_i(t')] + n_q(t)n_q(t') \tag{25}$$

y $P_e = P(y < 0)$.

La ecuación (25), que incluye productos de variadas gaussianas, recibe la designación de *forma cuadrática*. Se le puede simplificar por medio de un proceso de diagonalización al definir a cuatro nuevas variadas:

$$\alpha_i = A + \frac{n_i(t) + n_i(t')}{2} \qquad \alpha_q = \frac{n_q(t) + n_q(t')}{2}$$
$$\beta_i = \frac{n_i(t) - n_i(t')}{2} \qquad \beta_q = \frac{n_q(t) - n_q(t')}{2} \tag{26a}$$

ntonces, haciendo

$$\alpha = |\alpha_i + j\alpha_q| \qquad \beta = |\beta_i + j\beta_q| \tag{26b}$$

e puede demostrar que $y(t) = \alpha^2 - \beta^2$ y en consecuencia

$$P_e = P(y < 0) = P(\alpha^2 < \beta^2) = P(\beta > \alpha) \tag{27}$$

l último paso procede puesto que α y β son no negativas. Ahora se tiene una expre-ón idéntica a la de la manipulación no coherente por corrimiento de frecuencia con y β sustituyendo a y_1 y y_0. Además, como puede el lector comprobar, de las defini-ones, α es de Rice y β es de Rayleigh. La única diferencia aquí reside en que los tér-inos de ruido cuadráticos medios se han reducido en un factor de 2, por ejemplo, uponiendo que $n_q(t)$ y $n_q(t')$ son independientes, $\overline{\alpha_q^2} = (\overline{n_q^2} + \overline{n_q^2})/4 = N/2$. Por tanto,

$$P_e = \tfrac{1}{2}e^{-\rho} \tag{28}$$

omo se obtuvo de la ecuación (19) con ρ sustituida por $\rho/2$, lo que refleja la reduc-ón de ruido.

Con relación a las curvas de la figura 10.15, se aprecia que la manipulación por orrimiento de fase coherente diferencial tiene de 2 a 3 dB de ventaja sobre la detección o coherente y un precio de menos de 1 dB en comparación con la manipulación or corrimiento de fase coherente en $P_e \leq 10^{-4}$. Y recuérdese que la manipula-ón por corrimiento de fase coherente diferencial no requiere sincronización por se-arado. En forma adicional, con una ligera modificación de la onda transmitida, se eriva con facilidad una señal de temporización para el convertidor A/D. La única esventaja significativa es que dado el tiempo de retardo fijo $1/r$ en el detector, el stema se amarra en una velocidad de señalización específica, impidiéndose con ello transmisión a velocidad variable (asincrónica). Una incomodidad menor reside en hecho de los errores tienden a ocurrir en grupos de dos (¿por qué?).

Para la manipulación por corrimiento de fase coherente diferencial μ-aria, la robabilidad de error tiene la misma forma que en la ecuación (24), salvo que se de-incrementar a ρ en un factor de

$$\frac{\operatorname{sen}^2(\pi/\mu)}{2\operatorname{sen}^2(\pi/2\mu)} \tag{29}$$

ara obtener la misma P_e que en la detección coherente. Con $\mu = 4$ el precio es alre-edor de 2 dB y se incrementa en forma asintótica hacia los 3 dB conforme $\mu \to \infty$. sí, sólo para μ grande se hace que la detección diferencial sufra en forma completa referencia de fase ruidosa.

jemplo 10.3

a transmisión de datos a alta velocidad sobre canales telefónicos de voz ha sido un ma de interés práctico intenso durante muchos años. Uno de los diseños más re-

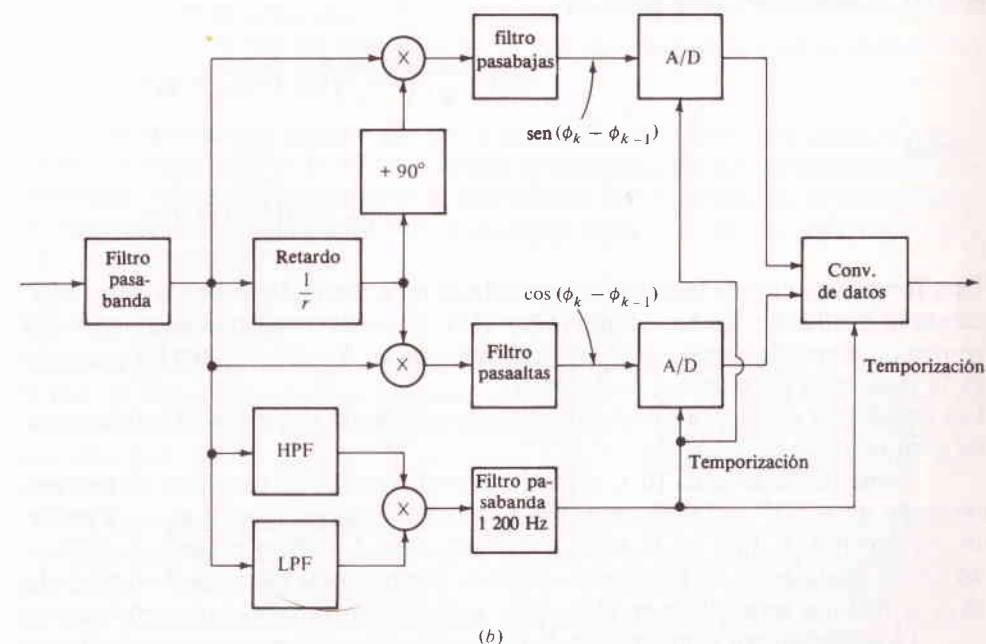
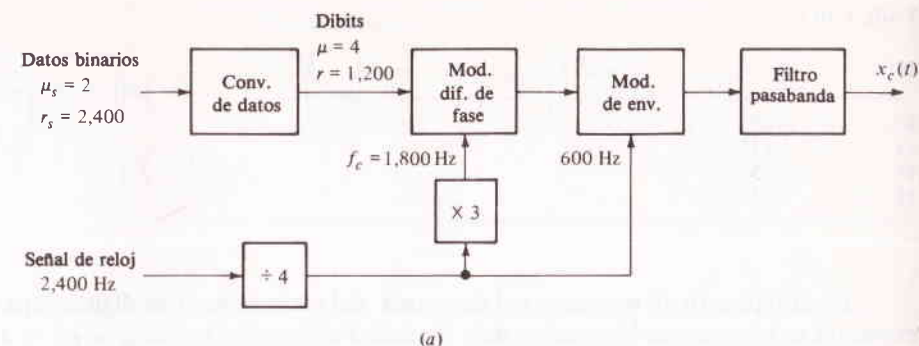



FIGURA 10.19
Sistema de manipulación por corrimiento de fase cuaternario. (*a*) Transmisor; (*b*) receptor

cientes y de más éxito, empleado en los equipos de datos del Sistema Bell modelos 201, 205 y 207, se describe aquí. Incorporando la modulación de manipulación por corrimiento de fase coherente diferencial cuaternaria, se obtiene la transmisión sincrónica a velocidades superiores a 2 400 bits/s en líneas telefónicas que han sido *acondicionadas* (igualadas) para señales digitales.

TABLA 10.1

Dibit	$\phi_k - \phi_{k-1}$	sen $(\phi_k - \phi_{k-1})$	cos $(\phi_k - \phi_{k-1})$
00	+45°	+	+
01	+135°	+	−
11	−135°	−	−
10	−45°	−	+

En la figura 10.19 se muestra el diagrama del transmisor. Los dígitos binarios ntrantes se agrupan en bloques de dos, llamados *dibits*, por lo que $\mu = \mu_s^2 = 4$ y $r = r_s/2 = 1\,200$. Los dibits modulan en forma diferencial a la fase de la portadora y a onda de manipulación por corrimiento de fase coherente diferencial después se nodula en envolvente para producir

$$x_c(t) = \sum_k p\left(t - \frac{k}{r}\right) \cos(2\pi f_c t + \phi_k)$$

donde

$$p(t) = \cos^2\left(\frac{\pi r t}{2}\right) \Pi(2rt)$$

Esta forma de pulso de banda base equivalente no es limitada en banda pero, aplicando la dualidad a las ecuaciones (17) y (18), se puede demostrar que $P(f)$ tiene contenido despreciable para $|f| > r$. En consecuencia, $B_T \approx 2r = 2\,400$ Hz centrado n la frecuencia portadora $f_c = 1\,800$ Hz. Las frecuencias portadora y de modulación de envolvente se derivan de la frecuencia le reloj entrante de 2 400 Hz.

Como indica la tabla 10.1, el corrimiento diferencial de fase tiene un término gregado de + 45° de tal manera que ϕ_K nunca puede ser igual a ϕ_{K-1} y existe n corrimiento de fase en la onda modulada cada $1/r$. Este aspecto, combinado con la modulación de envolvente, produce componentes discretos de frecuencia n $f_c \pm 600$ que se emplean en el receptor para generar la señal de temporización.

La detección por comparación de fase se efectúa como se muestra en la figura 0.19*b*, dando las salidas sen$(\phi_K - \phi_{K-1})$ y cos$(\phi_K - \phi_{k-1})$. Así, de acuerdo con a tabla 10.1, hay una correspondencia uno a uno entre las polaridades de salida y os dígitos binarios en cada dibit. El convertidor de datos interpola después los dos ígitos regenerados para producir una salida binaria sucesiva. La temporización se eriva de la mezcla de las salidas del filtro pasabajas y pasaaltas, los cuales cortan en 200 Hz, y seleccionan la frecuencia diferencia 1 800 − 600 = 1 200 Hz de las susochas componentes senoidales. Este procedimiento se emplea —en vez de filtraje— recto de la componente de 1 200 Hz, para compensar mejor la distorsión por retardo.

† Baker (1962) detalla las realizaciones de los componentes físicos para las funciones de transmisor y receptor, algunas de las cuales son muy ingeniosas.

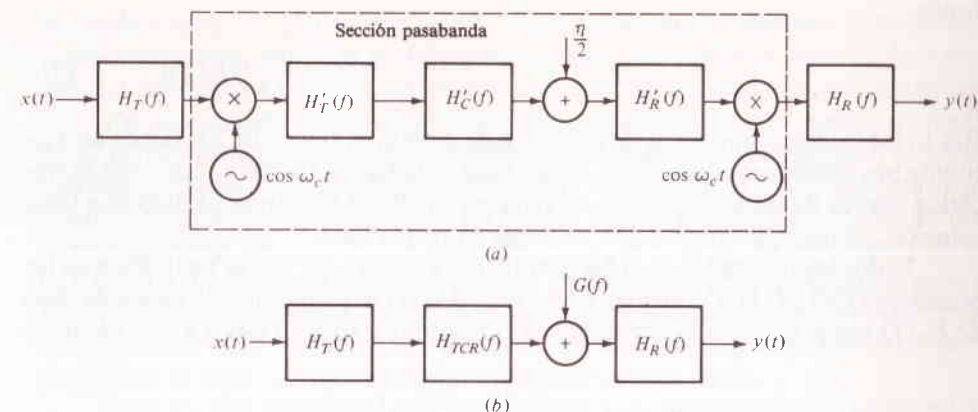



FIGURA 10.20
Modulación lineal coherente. (*a*) Diagrama a bloques de la porción analógica; (*b*) sistema de banda base equivalente.

Las pruebas han demostrado que la probabilidad de error del sistema es menor de 10^{-5} cuando $S/N = 15$ dB. Un sistema de manipulación por corrimiento de fase coherente diferencial completamente optimizado sin distorsión en la transmisión, sin interferencia intersimbólica, etc., daría lugar a una $P_e = 10^{-5}$ con alrededor de 3 dB menos potencia.

Modulación lineal coherente

Como un gran final del estudio de los métodos de modulación digital, se observará a la familia entera de la modulación lineal (AM, DSB, VSB, etc.) con detección coherente. Al estudiante que ha llegado hasta aquí, se le ruega tenga la deferencia de recordar que, por un cambio, se ha dejado para lo último uno de los casos más sencillos.

La porción de señal analógica de un sistema de modulación lineal coherente se puede apreciar en el diagrama de la figura 10.20*a*; las primas denotan unidades pasabanda. Suponiendo que el oscilador del receptor se encuentra sincronizado de modo perfecto y los filtros terminales de banda base no responden arriba de f_c, el análisis de la traslación de frecuencia demuestra que, en cuanto a la transmisión de la señal concierne, la sección pasabanda completa más el modulador y el demodulador es equivalente a un filtro *pasabajas*

$$H_{TCR}(f) = \tfrac{1}{4}[H'_{TCR}(f - f_c) + H'_{TCR}(f + f_c)] \qquad |f| < f_c \qquad (30a)$$

donde

$$H'_{TCR}(f) = H'_T(f)H'_C(f)H'_R(f) \qquad (30b)$$

Asimismo, el ruido equivalente pasabanda es gaussiano y tiene densidad espectral

$$G(f) = \tfrac{1}{4}[G'(f - f_c) + G'(f + f_c)] \qquad |f| < f_c \qquad (31a)$$

onde

$$G'(f) = \frac{\eta}{2} |H'_R(f)|^2 \qquad (31b)$$

Por lo tanto, para propósitos de análisis o de diseño, la figura 10.20*a* puede ser sustituida por *sistema equivalente de banda base* de la figura 10.20*b* y, dentro de la tradición clásica de las matemáticas, se reduce con ello el problema de algo que tenía solución, es decir, la transmisión de datos en banda base.

Todos los resultados de la sección 10.1 se aplican aquí con la sustitución de las ecuaciones (30) y (31). De manera específica, si $G(f)$ es plana sobre la banda de paso de $H_R(f)$ y si se suprime la portadora en el transmisor (o sea DSB o VSB), entonces

$$\begin{aligned} P_e &\geq 2\left(1 - \frac{1}{\mu}\right)Q\left[\sqrt{\frac{6\rho}{\mu^2 - 1}}\right] \\ &\approx \left(1 - \frac{1}{\mu}\right)\sqrt{\frac{\mu^2 - 1}{3\pi\rho}}\exp\left(-\frac{3\rho}{\mu^2 - 1}\right) \qquad \rho \gg 1 \end{aligned} \qquad (32)$$

en donde el límite inferior requiere de filtros terminales óptimos y es lo mismo que para transmisión óptima en banda base. Por lo tanto, la curva polar de banda base de la figura 10.15 también es válida para modulación lineal binaria con detección coherente y portadora suprimida. El caso de portadora *no suprimida* merece una especial atención aquí, puesto que con detección coherente, la potencia desperdiciada en la portadora no ofrece beneficios que compensen.

Por último, se debe notar, en el sistema equivalente de banda base que es posible la *igualación de banda base* directa con la modulación lineal coherente y corregirá la distorsión lineal introducida por el canal pasabanda. Tal igualación no funciona tan bien en los sistemas incoherentes porque la detección incoherente es un proceso no lineal y, en consecuencia, la distorsión lineal en el canal pasabanda produce distorsión no lineal en banda base.

EJERCICIO 10.4 Justifíquese ya sea la ecuación (30) o la (31), la que se encuentre más interesante. (*Sugerencias*: Para la ecuación (30), considérese una señal de banda limitada en $B < f_c$ aplicada al modulador; empléese el teorema de la traslación de frecuencia y sígase a través del sistema con unos bosquejos sencillos de los espectros. Para la ecuación (31), escríbase el ruido pasabanda en la entrada del detector en la forma de portadora en cuadratura, multiplíquesele por cos $\omega_c t$, y empléense los resultados de la Sec. 7.3.)

10.3 CODIFICACIÓN PARA CONTROL DE ERRORES

Se ha visto que la probabilidad de error en la transmisión digital es una función directa de ρ o en forma equivalente, de S/N. Si para un sistema dado, se limita la po-

tencia de la señal a algún valor máximo y los errores son, no obstante, frecuentes de modo inaceptable, entonces se debe buscar algún otro medio de mejorar la seguridad. La codificación para control de errores, a menudo constituye la mejor solución.

En pocas palabras, la codificación para control de errores es el empleo calculado de la *redundancia*. La teoría de la información sugiere que se agreguen en forma sistemática dígitos extra al mensaje transmitido, dígitos que en sí no conducen información, sino que hacen posible que el receptor detecte o incluso corrija los errores en los dígitos que llevan la información. En teoría, es posible una transmisión cercana a aquella sin errores; de manera práctica, existe el inevitable compromiso entre la seguridad en la transmisión, la eficiencia y la complejidad del equipo terminal. Considerando este factor, se ha ideado una gran variedad de códigos para la detección y corrección de errores para adecuarse a diferentes aplicaciones.

Esta sección constituye una introducción a la codificación para control de errores. Se tratará sólo con los códigos binarios, puesto que se ha hecho poco, relativamente, en el terreno de los códigos multinivel prácticos. (Recuérdese el aspecto exclusivo de los dígitos binarios: Si se sabe sólo cuáles dígitos son erróneos, inmediatamente se determinan los dígitos correctos.) Además, el procedimiento será en forma primordial cualitativo, dado que la teoría de la codificación ha evolucionado desde la magia negra de sus primeros días hasta llegar a ser una disciplina matemática muy compleja. En consecuencia, se está sólo rascando la superficie de un tema fascinante; el lector cuyo interés se haya despertado en relación con este tema y tenga aptitud para el álgebra moderna, encontrará material adicional en la lista de lectura suplementaria.

Conceptos y compromisos en la codificación

A modo de introducción, considérese un código repetido en el cual cada dígito de mensaje binario se repite tres veces —una burda analogía con la repetición de palabras cuando se está tratando de hablarle a alguien que está en el otro extremo de un salón ruidoso. Las palabras de código permitidas son entonces **000** y **111**, por lo que cualquier otra expresión recibida como **101** indica, a las claras, la presencia de errores. Para corregir errores sencillos, se podría emplear una decisión de regla mayoritaria y la siguiente tabla de decodificación:

Dígito decodificado	0	1
Palabras recibidas	000	111
	001	110
	010	101
	100	011

Esto produce un *código para corrección de error sencillo* y los errores por decodificación ocurren sólo cuando hay dos o tres dígitos erróneos en una palabra; por

ejemplo, dos errores cambian de **000** a **101**, etc. Por lo tanto, suponiendo que la *probabilidad de error por dígito* es ϵ, se encuentra la probabilidad de error por decodificación de la distribución binomial, ecuación (1), (Sec. 3.4), como

$$P_e = P\,(2 \text{ o } 3 \text{ errores en 3 dígitos}) = P_3(2) + P_3(3)$$
$$= \binom{3}{2}\epsilon^2(1-\epsilon) + \binom{3}{3}\epsilon^3 = 3\epsilon^2 - 2\epsilon^3 \qquad (1)$$

Puesto que ϵ es la probabilidad de error sin codificación, y como $\epsilon < ½$ en cualquier canal razonable, la codificación ha mejorado en realidad la seguridad.

El código de triple repetición, también funciona en la *detección de error doble* si se prescinde de la corrección de error sencillo; es decir, que a cualquier palabra de código recibida diferente de **000** o **111** se le trata como un error detectado, pero no corregido. Los errores por decodificación, en el sentido de errores no detectados, ocurren con probabilidad

$$P_e = P_3(3) = \epsilon^3 \qquad (2)$$

la cual es obviamente más pequeña que en la ecuación (1). No obstante la trivialidad de este código de ejemplo, conduce a tres conclusiones generales e importantes acerca de la codificación para control de error.

1 Agregando los dígitos extra, a los que se les designa *dígitos de comprobación*, las palabras de código pueden ser "muy diferentes" entre sí. En forma analítica, la diferencia entre cualesquier dos palabras binarias se mide en términos de la *distancia Hamming d*, definida en forma sencilla como el número de lugares en los que las palabras tienen dígitos diferentes; así, toma d errores (en los lugares de la derecha) para cambiar de una palabra a otra. Continuando con esta idea, un código que detecta o corrige a K errores por palabra debe consistir de palabras de código que tengan

$$d_{\text{mín}} = \begin{cases} K+1 & \text{detección de error} \\ 2K+1 & \text{corrección de error} \end{cases} \qquad (3)$$

una aseveración cuyo análisis se deja al estudiante. El código de repetición triple obviamente tiene $d = 3$ por lo que, como ya se vio, puede detectar $K = 2$ errores o corregir $K = 1$ error por palabra. Para un código arbitrario con

k = dígitos de mensaje por palabra
q = dígitos de comprobación por palabra
$n = k + q$ = dígitos totales por palabra

hay 2^k palabras de código binario (formadas con k dígitos de mensaje) fuera de $2^n = 2^q \times 2^k$ palabras de n-dígitos posibles. De conformidad, los dígitos de comprobación se deben escoger de tal forma que las 2^k palabras de código satisfagan los requisitos de distancia (3).

2 Si la probabilidad de error por dígito ϵ es razonablemente pequeña, entonces la probabilidad de $M + 1$ errores en una palabra de n-dígitos será mucho menor que la probabilidad de M errores, es decir, $P_n(M+1) \ll P_n(M)$. Para recalcar este punto, y como referencia, en la tabla 10.2 se listan expresiones aproximadas para $P_n(M)$ obtenidas por medio de desarrollos en serie binomial. Por lo tanto, si un código corrige o detecta K errores, la probabilidad de error por decodificación por palabra es

$$P_{e,\text{palabra}} = \sum_{i=K+1}^{n} P_n(i) \approx P_n(K+1) \qquad (4a)$$

siendo la aproximación bastante precisa si $n\epsilon \leq 0.1$. Como la mayor parte de los errores de decodificación se deben a $K + 1$ errores de dígito, de los cuales la fracción k/n son dígitos erróneos de mensaje (el resto son errores de dígitos de comprobación), la probabilidad de error neta por dígito de mensaje o por bit es

$$P_{e,\text{bit}} \approx \frac{k}{n}(K+1)\,P_{e,\text{palabra}} \qquad (4b)$$

TABLA 10.2 APROXIMACIONES EN SERIE PARA LA DISTRIBUCIÓN BINOMIAL

M	$P_n(M) = \binom{n}{M}\epsilon^M(1-\epsilon)^{n-M}$
0	$1 - n\epsilon + ½n(n-1)\epsilon^2 - ⅙n(n-1)(n-2)\epsilon^3$
1	$n\epsilon - n(n-1)\epsilon^2 + ½n(n-1)(n-2)\epsilon^3$
2	$½n(n-1)\epsilon^2 - ½n(n-1)(n-2)\epsilon^3$
3	$⅙n(n-1)(n-2)\epsilon^3$

3 La inserción de los dígitos de comprobación para control de error reducen la velocidad efectiva a que se transmiten los dígitos de mensaje. En forma cuantitativa, se define el *factor de eficiencia en la velocidad* de un código como†

$$\mathscr{E} = \frac{k}{k+q} = \frac{k}{n} \qquad (5)$$

por lo que si la velocidad burda de señalización es r, la velocidad de dígitos de mensaje es

$$r_m = \mathscr{E} r \qquad (6)$$

En términos generales, los códigos que están instrumentados con facilidad y además son efectivos en el control de errores, requieren relativamente grande porcentaje de dígitos de comprobación. Así, el control práctico de errores tiende a ir de la mano con la reducción de velocidad de bits.

† No se debe confundir con la eficiencia de la comunicación definida en la ecuación (8), (Sec. 9.4).

Pero hay otras formas menos elegantes de disminuir los errores a expensas de la velocidad de señalización, permaneciendo fija la potencia de la señal. Y para evaluar en forma apropiada los méritos de un código dado, se debe considerar por lo menos otra opción, a saber, la reducción de la velocidad de señalización sin codificación. Reduciendo a r, se incrementa a $p = S_R/\eta r$ (si los filtros terminales se ajustan de conformidad) y con ello se disminuye la probabilidad de error.

Supóngase, para efectos de comparación, que se propone un código determinado con eficiencia de velocidad $\mathscr{E}$ para su empleo en un canal binario en banda base que tenga

$$\epsilon = Q(\sqrt{2\rho}) \qquad \rho = \frac{S_R}{\eta r}$$

de tal manera que $P_{e,\,\mathrm{bit}}$ la da la ecuación (4*b*) con el valor anterior de ϵ. Por otra parte, se podría reducir simplemente la velocidad de señalización en un factor de $\mathscr{E}$, dando la misma velocidad de bits de mensaje con

$$P_{e\text{, no codificada}} = Q\left(\sqrt{\frac{2\rho}{\mathscr{E}}}\right) \tag{7}$$

Si la ecuación (7) es del mismo orden de magnitud que $P_{e,\,\mathrm{bit}}$ el valor del código particular sometido a consideración es cuestionable.

Ejemplo 10.4

El código de repetición triple tiene $k = 1$, $q = 2$, $n = 3$, y $\mathscr{E} = 1/3$. Si se emplea para corrección de error sencillo en un canal de banda base con $\rho = 7$ y $r = 1\,200$, la velocidad de bits de mensaje es $r_m = 1\,200/3 = 400$ y $\epsilon = Q(\sqrt{2P}) \approx 10^{-4}$ por lo que, de la ecuación (1),

$$P_{e,\mathrm{bit}} \approx 3\epsilon^2 \approx 3 \times 10^{-8}$$

La ecuación (4*b*) no se aplica en este caso, puesto que el dígito del mensaje decodificado siempre es erróneo cuando hay dos o tres errores.

Sin embargo, si se reduce la velocidad de señalización a $r = 300$ y no se emplea la codificación,

$$P_{e\text{, no codificada}} = Q\left(\sqrt{\frac{2\rho}{\mathscr{E}}}\right) = Q(\sqrt{42}) \approx 5 \times 10^{-9}$$

por lo que la sencilla reducción de velocidad en la señalización es superior a este código rudimentario.

Detección de error por comprobación de paridad

Para muchas aplicaciones los errores se pueden reducir sin alterarlos si sólo *se detectan* sin intentar una corrección inmediata. Esto es cierto, por ejemplo, en la telemetría de datos, cuando se reúne un gran número de valores para análisis estadístico; los valores erróneos, si se detectan, simplemente se omiten en el procesamiento posterior, y la pérdida es despreciable. De manera similar, dado un enlace de comunicación de dos vías, por el hecho de que se haya detectado un error se puede regresar al transmisor por medio de una acción apropiada, o sea, la retransmisión. Tal decisión de realimentación es ventajosa de manera especial si el sistema está sujeto a condiciones de transmisión variables. Cuando las condiciones son buenas y los errores no son frecuentes, resulta satisfactorio un código de baja redundancia con alta velocidad de datos; cuando las condiciones son desfavorables, como lo puede indicar una detección frecuente de errores, el transmisor puede conmutar a un código de más alta redundancia o cesar la transmisión en forma temporal. Pero con o sin realimentación, la detección de error sencilla es suficiente sólo si ϵ es pequeña y la probabilidad de los errores no detectados está en un adecuado bajo nivel.

Muchos códigos para detección de errores están basados en la noción de *paridad*. Se dice que la paridad de una palabra binaria es para cuando la palabra incluye un número par de **1**s, en tanto que la paridad impar significa un número impar de **1**s. Para la detección de errores por medio de la comprobación de paridad, se divide el mensaje en grupos de k dígitos y se agrega un dígito de comprobación a cada grupo tal que cada palabra de $(k + 1)$ dígitos tenga la misma paridad, digamos par. Así, el dígito de comprobación queda relacionado con los dígitos de mensaje por

$$c = m_1 \oplus m_2 \oplus \cdots \oplus m_k \tag{8a}$$

en donde se establece a $\oplus$ para significar ***adición del módulo 2***. La aritmética del módulo 2, definida en los dígitos binarios **0** y **1**, es igual que la aritmética ordinaria excepto en que $\mathbf{1} \oplus \mathbf{1} = \mathbf{0}$ y no existe diferencia entre la suma y la resta.† En consecuencia, la ecuación (8*a*) es equivalente a

$$m_1 \oplus m_2 \oplus \cdots \oplus m_k \oplus c = \mathbf{0} \tag{8b}$$

El factor de eficiencia es

$$\mathscr{E} = \frac{k}{k+1} \tag{9}$$

indicando una eficiencia razonable si k es grande.

De las 2^{k+1} palabras binarias posibles que tienen $k + 1$ dígitos, la codificación por comprobación de paridad excluye precisamente a la mitad, la mitad con paridad impar, asegurando con ello que el código tenga una distancia de Hamming $d \geq 2$ como se requiere para la detección de error sencillo. Por lo tanto, si la paridad de una palabra recibida es impar, se sabe que hay un error —o tres errores, o, en general, un número impar de errores. La detección de error se puede entonces implementar comprobando la paridad de cada palabra conforme ésta llega. Por supuesto que no

† En términos de componentes físicos, $\oplus$ es una compuerta EXLUSIVA-O.

es posible la corrección de error, dado que no se sabe *en dónde* se localizan los errores dentro de la palabra. Además, un número par de errores mantiene válida la paridad y, en consecuencia, pasa sin detectarse.

Pasando por alto todo excepto el caso de error doble, $P_{e,\text{ palabra}} \approx P_{k+1}(2) \approx \frac{1}{2}(k+1)k\epsilon^2$ y por ende, de la ecuación (4*b*),

$$P_{e,\text{bit}} \approx \frac{k}{k+1}\, 2P_{e,\text{ palabra}} \approx k^2\epsilon^2 \qquad (10)$$

Por ejemplo, si $k = 9$ y $\epsilon = 10^{-3}$, la codificación por comprobación de paridad abate la probabilidad de error en más de un orden de magnitud con una reducción de velocidad de sólo 9/10.

La probabilidad de un error detectado es también de interés porque indica la cantidad de datos que deben ser retransmitidos o descartados. Puesto que principalmente se detectan errores sencillos, se tiene *por palabra*

$$P_{de} \approx \binom{k+1}{1}\epsilon(1-\epsilon)^k \approx (k+1)\epsilon \qquad (11a)$$

En un mensaje de $N \gg 1$ dígitos totales de mensaje, hay N/k palabras, de las cuales $(N/k)(k+1)\epsilon$ tienen errores detectados. Si se descartan los errores detectados, el número fraccionario de dígitos de mensaje desechados es

$$\frac{1}{N}\left[k\,\frac{N}{k}(k+1)\epsilon\right] = (k+1)\epsilon = P_{de} \qquad (11b)$$

Si $k = 9$ y $\epsilon = 10^{-3}$, $P_{de} \approx 0.01$ o 1%.

Dos comentarios finales en relación con cuestiones prácticas: Primero, es preferible emplear la paridad de palabra impar y un número impar de dígitos de mensaje por palabra; esto asegura que cada palabra tenga por lo menos una transición, ayudando con ello a la sincronización y previniendo la pérdida aparente de la señal si el mensaje contiene una hilera extensa de semejantes dígitos. Segundo, como resultado del ruido de impulsos sobre circuitos conmutados o del desvanecimiento de corta duración en trayectorias de radio, los errores pueden tener la tendencia a ocurrir en *ráfagas* de varios dígitos sucesivos; dado que los errores múltiples arruinan la comprobación de paridad, los dígitos de comprobación deben *entrelazarse* de tal manera que los dígitos comprobados estén espaciados en forma amplia. Un ejemplo del entrelazado se da en la figura 10.21, en donde se indica una palabra de paridad por medio de líneas que conectan a los dígitos.

EJERCICIO 10.5 Lístense todas las palabras de código para un código de comprobación de paridad con $k = 3$ y verifíquese que $d \geq 2$.

FIGURA 10.21
Comprobación de paridad entrelazada; m = dígito de mensaje, c = dígito de comprobación.

Códigos de bloques para corrección de error

La idea de los códigos *de corrección de error* por cierto que insta e incita en mayor medida que la simple detección de error, sugiriendo que con ella se cumple la transmisión unilateral sin errores, cuya hipótesis estableció Shannon. Y hay numerosas aplicaciones, de manera notable los enlaces unilaterales, en donde la corrección de errores constituye una necesidad. Resulta que la codificación en el transmisor no es muy diferente de la detección o la corrección; el decodificador de recepción es el problema en los sistemas de corrección de errores. En consecuencia, cuando el equipo de transmisión tiene restricciones, por ejemplo, en satélites, debe de demostrar ser más práctico para corregir errores en el receptor que presentar facilidades de retransmisión, aún cuando se disponga de una trayectoria bilateral.

La codificación por comprobación de paridad se extiende con facilidad a la corrección de errores observando que la corrección requiere de la detección de un error y de su *ubicación* en la palabra. Así, si se hacen dos comprobaciones en la misma palabra pero en dos modelos diferentes, los errores ocasionan síntomas característicos de invalidación de la paridad.

Muchos códigos de bloque para la corrección de errores tienen palabras de n-dígitos en las cuales los primeros k dígitos son dígitos de mensaje y los restantes $q = n - k$ son dígitos de comprobación de paridad. A tal código se le designa ***código sistemático de comprobación de paridad (n, k)***, y el dígito i-ésimo de una palabra de código es

$$x_i = \begin{cases} m_i & i = 1, 2, \ldots, k \\ c_{i-k} & i = k+1, \ldots, n \end{cases} \qquad (12)$$

Por conveniencia analítica, una palabra típica de código se representa por medio de una matriz $n \times 1$ (o vector columna) $\mathbf{x}$. Por ejemplo, una palabra de código (7,4) sería

$$\mathbf{x} = [m_1 \quad m_2 \quad m_3 \quad m_4 \quad c_1 \quad c_2 \quad c_3]^T$$

la cual se ha escrito como la traspuesta de una matriz $1 \times n$ para ahorrar espacio.

Dados los dígitos de mensaje para una palabra en particular, los dígitos de comprobación se escogen de tal manera que

$$\mathbf{Hx} = \mathbf{0} \qquad (13)$$

donde la *matriz de comprobación de paridad* $\mathbf{H}$ es una matriz rectangular $q \times n$ de la forma

$$\mathbf{H} = \left.\begin{bmatrix} \overbrace{h_{11} & h_{12} & \ldots & h_{1k}}^{k} & \overbrace{\mathbf{1} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0}}^{q} \\ h_{21} & h_{22} & \ldots & h_{2k} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \ldots & \mathbf{0} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\ h_{q1} & h_{q2} & \ldots & h_{qk} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{1} \end{bmatrix}\right\} q \qquad (14)$$

Nótese que la porción derecha del segundo miembro de **H** es una matriz unitaria $q \times q$; por lo que, combinando las ecuaciones (12) y (14), cada fila de la ecuación (13) es una ecuación de comprobación de paridad que comprende sólo un dígito de comprobación, es decir,

$$h_{j1}m_1 \oplus h_{j2}m_2 \oplus \ldots \oplus h_{jk}m_k \oplus c_j = 0 \qquad j = 1, 2, \ldots, q$$

Los componentes h_{11}, h_{12}, etc., de **H** son dígitos binarios, pero no se les ha especificado aún dado que éstos determinan las propiedades de corrección de error del código para lo cual se precisa de más información.

Supóngase que se transmite una palabra **x** y, debido a los errores, resulta en la palabra recibida

$$\mathbf{y} = \mathbf{x} \oplus \mathbf{e} \qquad (15a)$$

en lo cual se conviene que para la suma dígito a dígito

$$y_i = x_i \oplus e_i \qquad i = 1, 2, \ldots, n \qquad (15b)$$

con **e** como el *modelo de error*, o sea,

$$e_i = \begin{cases} 1 & y_i \neq x_i \\ 0 & y_i = x_i \end{cases}$$

Como aclaración, si

$$\mathbf{x} = [0 \quad 1 \quad 0 \quad 0 \quad 1]^T$$

y

$$\mathbf{y} = [0 \quad 1 \quad 1 \quad 0 \quad 0]^T$$

entonces

$$\mathbf{e} = [0 \quad 0 \quad 1 \quad 0 \quad 1]^T$$

lo que indica que hay errores en el tercero y quinto dígitos.

Si se pudiera determinar a **e** en el receptor, se corregirían todos los errores; pero el encontrar a **e** de **y** requiere que se conozca a **x**, ¡la palabra transmitida! Sin embargo, se puede calcular un *síndrome* de q-dígitos

$$\mathbf{s} = \mathbf{H}\mathbf{y} \qquad (16a)$$

lo cual proporciona alguna información útil. De manera específica, de las ecuaciones (13) y (15),

$$\mathbf{s} = \mathbf{H}\mathbf{x} \oplus \mathbf{H}\mathbf{e} = \mathbf{H}\mathbf{e} \qquad (16b)$$

por lo que $\mathbf{s} = \mathbf{0}$ cuando $\mathbf{e} = \mathbf{0}$, es decir, sin errores. Además, si **y** tiene sólo un error, digamos en el dígito de mensaje j-ésimo,

$$\mathbf{s} = [h_{1j} \quad h_{2j} \quad \ldots \quad h_{qj}]^T \qquad j \leq k \qquad (17)$$

y éste es idéntico a la columna j-ésima de la matriz de comprobación de paridad. Por lo tanto, el síndrome indica sin ambigüedades la condición de que no hay error o la posición de un error único, siempre y cuando *todas las columnas de* **H** *sean diferentes y diferentes de cero*; y entonces se tiene un código de corrección de error único.

Cuando se emplea de este modo, un código (n, k) tiene $P_{e,\,\text{palabra}} \approx P_n(2)$ y

$$P_{e,\text{bit}} \approx \frac{k}{n} 2P_n(2) \approx k(n-1)\epsilon^2 \qquad (18)$$

Los errores múltiples ocasionan complicaciones, sin embargo; dado que se pueden omitir los errores reales y de manera equivocada "corregir" otro dígito de mensaje, empeorando las cosas. En consecuencia, a menos que ϵ sea tan pequeña, como para que los errores múltiples sean muy raros, resulta deseable un código más contundente.

Por desgracia, idear un síndrome apropiado y una matriz de comprobación de paridad para la corrección de errores múltiples es una tarea mucho más complicada —tanto, que los primeros códigos para corrección de error doble se crearon por inspiración en el trabajo de prueba y error, más que por el diseño de un método específico. Finalmente, Slepian (1956) situó la teoría de la codificación sobre sólidos fundamentos matemáticos cuando descubrió su relación con los conceptos del álgebra moderna. Poco después, empleando la teoría de los campos de Galois,† Bose, Chaudhuri y Hocquenghem desarrollaron una categoría de códigos para corrección de errores múltiples (ahora se les designa códigos BCH) que son eficientes y presentan requisitos en sus componentes físicos de manera relativa sencillos para la codificación y decodificación. No es necesario decir que tales códigos quedan fuera del alcance de este texto.

EJERCICIO 10.6 Verifíquese que un código de triple repetición es un código (3, 1) sistemático de comprobación de paridad con

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

En particular, demuéstrese que $\mathbf{H}\mathbf{x} = \mathbf{0}$ y encuéntrese a **s** para posición de error único.

Ejemplo 10.5 Códigos de Hamming

Antes del descubrimiento de Slepian e incluso antes de la formulación matricial de los códigos de bloques, Hamming (1950) ideó una clase más elegante de códigos de bloques. En esta notación, y tratándose sólo de la corrección de error sencillo, la estrategia de Hamming es como sigue: Si hay q dígitos de comprobación por palabra, entonces el síndrome es una palabra de q dígitos que se puede hacer para des-

† Se ha corrido el rumor de que un matemático teórico que se especializaba en los campos de Galois, abandonó el tema hasta que se les encuentre aplicación práctica.

cifrar en forma binaria y situar así la posición exacta de un error sencillo, si lo hay. Con $q = 3$, por ejemplo, s = **000** significa "que no hay error" (como antes), s = **001** significa "error en el primer dígito", etc.

Dado que se requieren indicaciones de $n + 1$ errores ("sin error" o de un error en cualquiera de los n dígitos de palabra de código), y puesto que hay 2^q palabras de síndrome diferentes, los números de dígitos de comprobación y dígitos de mensaje en un código de Hamming están relacionados por medio de

$$2^q = k + q + 1 \qquad \text{donde } k + q = n \tag{19}$$

De conformidad, el factor de eficiencia es

$$\mathscr{E} = \frac{k}{n} = 1 - \frac{1}{n}\log_2(n+1) \tag{20}$$

por lo que se obtiene una eficiencia razonable empleando palabras de código largas.

La matriz de comprobación de paridad se construye con facilidad aprovechando lo estipulado con anterioridad para s y el hecho de que s es igual a la columna j-ésima de **H** cuando hay un error único en el dígito j-ésimo. Por tanto, leyendo de izquierda a derecha, las columnas de **H** son simplemente las versiones binarias de los números 1, 2, . . . , n, como se muestra a continuación para un código de Hamming (7, 4).

$$\mathbf{H} = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \tag{21a}$$

Entonces, por comparación con la ecuación (14) que no es un código sistemático; dado que las posiciones de los dígitos de comprobación deben corresponder a las columnas de **H** que tienen sólo un **1**, se desprende que la palabra de código (7, 4) tiene la forma

$$\mathbf{x} = [c_1 \quad c_2 \quad m_1 \quad c_3 \quad m_2 \quad m_3 \quad m_4]^T \tag{21b}$$

y las ecuaciones para los dígitos de comprobación son

$$\begin{aligned} c_1 &= m_1 \oplus m_2 \oplus m_4 \\ c_2 &= m_1 \oplus m_3 \oplus m_4 \\ c_3 &= m_2 \oplus m_3 \oplus m_4 \end{aligned} \tag{21c}$$

Obsérvese que cada dígito de mensaje está comprobado con por lo menos dos dígitos de comprobación, lo cual es esencial en la corrección de errores.

EJERCICIO 10.7 Tomando los parámetros de sistema del ejemplo 10.4, encuéntrese a $P_{e,\,\text{bit}}$ para un código de Hamming (7, 4), y compárese con P_e, no codificada con $\mathscr{E} = 4/7$. *Respuesta*: $P_{e,\,\text{bit}} \approx 2 \times 10^{-7}$, $P_{e,\,\text{no codificada}} \approx 4 \times 10^{-7}$.

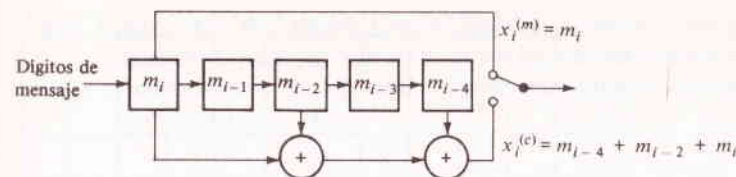



FIGURA 10.22
Codificación convolucional empleando registro de corrimiento.

Códigos convolucionales

Estos códigos, también conocidos como secuenciales o recurrentes, difieren de los códigos de bloque en que los dígitos de comprobación se intercalan en forma continua en la corriente codificada de bits en vez de agruparse en palabras. Por lo tanto, el procedimiento de codificación/decodificación es un proceso continuo, eliminando los componentes físicos de separación o almacenamiento que se requieren en los códigos de bloque. La teoría de los códigos convolucionales es bastante intrincada, pero se puede demostrar el principio utilizando un ejemplo sencillo.

La figura 10.22 es un codificador convolucional; formado de un *registro de corrimiento* de cinco celdas, a través del cual se mueven los dígitos de izquierda a derecha, más un sumador de módulos 2 y un conmutador. El conmutador recoge en forma alterna al dígito de mensaje i-ésimo $x_i^{(m)} = m_i$ y al dígito de comprobación i-ésimo

$$\begin{aligned} x_i^{(c)} &= m_{i-4} \oplus m_{i-2} \oplus m_i \\ &= x_{i-4}^{(m)} \oplus x_{i-2}^{(m)} \oplus x_i^{(m)} \end{aligned} \tag{22}$$

Los dígitos de mensaje se desplazan después sobre una celda y se repite el proceso. Por consiguiente, la secuencia transmitida $x_1^{(m)} x_1^{(c)} x_2^{(m)} x_2^{(c)} \dots$ tiene dos veces la velocidad de bits de los datos que llegan, y $\mathscr{E} = 1/2$. Esta alta redundancia no es una necesidad, pero muchos códigos convolucionales tienen relativamente bajos factores de eficiencia en velocidad.

En el decodificador se forma el dígito de síndrome i-ésimo, de la secuencia recibida, de conformidad con la regla

$$s_i = y_{i-4}^{(m)} \oplus y_{i-2}^{(m)} \oplus y_i^{(m)} \oplus y_i^{(c)} \qquad i \geq 5$$

donde, debido a los errores, $y_i^{(m)} = x_i^{(m)} \oplus e_i^{(m)}$, etc. Así, se desprende de la ecuación (22) que

$$s_i = e_{i-4}^{(m)} \oplus e_{i-2}^{(m)} \oplus e_i^{(m)} \oplus e_i^{(c)} \qquad i \geq 5 \tag{23a}$$

	$e_1^{(m)}$	$e_1^{(c)}$	$e_2^{(m)}$	$e_2^{(c)}$	$e_3^{(m)}$	$e_3^{(c)}$	$e_4^{(m)}$	$e_4^{(c)}$	$e_5^{(m)}$	$e_5^{(c)}$	$e_6^{(m)}$	$e_6^{(c)}$
s_1	X	X										
s_2			X	X								
s_3	X				X	X						
s_4			X				X	X				
s_5	X				X				X	X		
s_6			X				X				X	X

FIGURA 10.23

lo cual comprueba la paridad en el sentido de que $s_i = 1$ si hay un número impar de errores, mientras que de otra manera $s_i = 0$. Para el transitorio de puesta en marcha, $1 \leq i \leq 4$,

$$s_1 = e_1^{(m)} \oplus e_1^{(c)} \qquad s_3 = e_1^{(m)} \oplus e_3^{(m)} \oplus e_3^{(c)}$$
$$s_2 = e_2^{(m)} \oplus e_2^{(c)} \qquad s_4 = e_2^{(m)} \oplus e_4^{(m)} \oplus e_4^{(c)} \tag{23b}$$

En la figura 10.23 se presenta la ecuación (23) en forma gráfica; por ejemplo, los xs en la primera columna indican que $e_i^{(m)}$ aparece en s_1, s_3 y s_5. El estudio de esta figura revela que si hay dos o tres **1**s en $s_1s_3s_5$, entonces más probablemente $e_1^{(m)} = 1$, significa que $y_1^{(m)}$ es errónea y se debe corregir; de manera similar, se debe corregir a $y^{(m)}$ si hay más **1S** que **0S** en $s_2s_4s_6$, y así sucesivamente. El estudiante puede verificar que este algoritmo conocido como *decodificación de umbral*, corregirá a cuatro errores sucesivos (incluidos los dígitos de comprobación) siempre y cuando los ocho dígitos siguientes estén libres de error.

La decodificación de umbral es efectiva en forma particular para aquellos canales donde el principal problema sean las *ráfagas de errores* aisladas debidas al ruido de impulsos. Otro algoritmo de decodificación para códigos convolucionales es el método probabilístico o *secuencial*, inventado por Wozencraft. Dado que la teoría de los códigos convolucionales no está tan bien desarrollada como la de los códigos de bloque, es difícil hacer una evaluación precisa de sus méritos relativos.

10.4 PROBLEMAS

10.1 (Secc. 10.1) Considérese un sistema que tenga $p(t) = \text{senc}\, r(1 + \epsilon)t$, $0 < \epsilon \ll 1$, por lo que la velocidad de señalización es ligeramente menor que la velocidad sincrónica $r(1 + \epsilon)$. Los datos constituyen una secuencia de $2M + 1$ alternando 1s y 0s, es decir, $A_k = (-1)^k A$, $k = 0, \pm 1, \pm 2, \ldots, \pm M$. Demuéstrese que el término de interferencia intersimbólica de la ecuación (3) en $m = 0$ es $[2A\epsilon/(1 + \epsilon)]\, \Sigma^M{}_k = 1$ senc $k\,\epsilon \approx 2M\,\epsilon A$ si $M\epsilon \ll 1$.

10.2 (Secc. 10.1) Un sistema binario experimenta interferencia intersimbólica tal que, en ausencia de ruido, los valores de $y(tm)$ y sus probabilidades son como se tabula abajo, cuando se envía un **1**; la tabla se aplica también cuando envía un **0**, excepto que se sustituye a A por $-A$.

$y(t_m)$:	$A - \alpha$	A	$A + \alpha$
$P[y(t_m)]$:	¼	½	¼

(*a*) Suponiendo ruido gaussiano, obténgase una expresión para P_e en términos de A, α, y σ.

(*b*) Evalúese a P_e para $A/\sigma = 4.0$ cuando $\alpha/A = 0.05$ y 0.25. Compárese con P_e cuando $\alpha = 0$.

10.3 (Secc. 10.1) Una señal binaria polar se envía por dos canales ideales diferentes, que se denotan por a y b, al mismo destino, donde las señales se combinan como se indica en la figura P10.1. Este arreglo se conoce como sistema de *diversidad*. El canal a tiene un amplificador con una ganancia de voltaje ajustable K, y n_b son variadas gaussianas independientes con $\overline{n_a^2}\ 3\ \Phi_a^2$, $\overline{n_b^2} = \phi_b^2$.

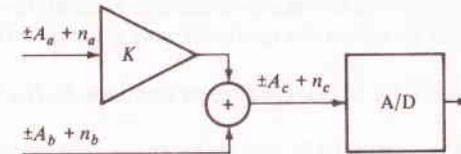



FIGURA P10.1

(*a*) Encuéntrese el valor de K que aumente al máximo a A_c/σ_c y con ello se reduzca al mínimo a P_e.
Respuesta: $K_{\text{ópt}} = A_a\sigma_b^2/A_b\sigma_a^2$.

(*b*) Tomando $K = K_{\text{ópt}}$, $A_b/\sigma_b = 3.0$, y $A_a/\sigma_a = \alpha(A_b/\sigma_b)$, constrúyase gráfica de P_e para el sistema de diversidad y para el canal a sólo como una función de α, $0 \leq \alpha \leq 2$.

10.4 (Secc. 10.1) Sea $\pm\, a\Pi[(t - \tau/2)/\tau]$ más el ruido blanco $n(t)$ la entrada a un filtro ideal de integrado y descarga. Demuéstrese que la salida en $t = \tau$ tiene $(A/\phi)^2 = 2E_R/\eta$, por lo que el comportamiento es igual al de filtraje acoplado. (*Sugerencia*: Empléese $\phi^2 = \mathbf{E}\{[\int_0^\tau n(t)dt]^2\}$ y recuérdese que $\mathbf{E}[n(t_1)n(t_2)] = R_n(t_1 - t_2) = (\eta/2)\delta(t_1 - t_2)$.)

10.5 (Secc. 10.1) Considérese un sistema binario polar en la forma de la figura 10.1 con $a_k = \pm\, a$, $H_T(f) = 1$, $G(f) = \eta/2$ y $H_R(f) = P_s^*(f)H_c^*(f)$, donde $P_s(f) = \mathscr{F}[P_s(t)]$, por lo que el filtro de recepción está acoplado a la forma del pulso recibido. Como este arreglo no conforma al pulso de salida $p(t)$ por consideraciones de interferencia intersimbólica, sólo se emplea cuando $p_s(t)$ es limitado en tiempo (pero no necesariamente rectangular) con duración mucho menor que $1/r$.

(*a*) Considerando la transmisión de un pulso aislado, demuéstrese que $P(f)$ es real por lo que $A = ap(0) = a\int_{-\infty}^{\infty} P(f)df$, y por ello

$$\left(\frac{A}{\sigma}\right)^2 = \left[\frac{\int_{-\infty}^{\infty}|P_s(f)H_c(f)|^2 df}{\int_{-\infty}^{\infty}|P_s(f)|^2 df}\right]\left(\frac{2S_T}{\eta r}\right)$$

donde $S_T = rE_T$ y E_T es la energía del pulso transmitido.
(*b*) Compruébese que esto es equivalente a la ecuación (14).

10.6 (Secc. 10.1) Empléense las ecuaciones (15*a*) y (16*a*) para encontrar y bosquejar a $P(f)$ y a $p(t)$ para:
(*a*) $P_\beta(f) = (\frac{1}{2}\beta)\Pi(f/2\beta)$, $\beta = r/4$
(*b*) $P_\beta(f) = (1/\beta)\Lambda(f/\beta)$, $\beta = r/2$

10.7 (Secc. 10.1) Desarróllense todos los detalles que conducen a las ecuaciones (17*b*) y (18*a*) partiendo de la ecuación (17*a*).

10.8★ (Secc. 10.1) Considérese un sistema binario polar con $r = 2 \times 10^4$, $ps(t) = \Pi(2rt)$, $|H_C(f)| = 10^{-2}$, $G(f) = 10^{-10}(1 + 3 \times 10^{-4}|f|)^2$, y $p(t)$ por la ecuación (18*b*) por lo que $P(f) = (1/r)\cos^2(\pi f/2r)\Pi(f/2r)$.
(*a*) Encuéntrese y bosquéjese la relación de amplitudes para los filtros terminales óptimos.
(*b*) Calcúlese el valor de S_T que se necesita para que $P_e = 10^{-6}$.
Respuesta: $S_T = 3.49$.

10.9★ (Secc. 10.1) Repítase el problema 10.8 para un sistema cuaternario ($\mu = 4$) con $r = 100$, $p_s(t) = \Pi(10rt)$, $|H_C(f)| = 10^{-3}/(1 + 32 \times 10^{-4}f^2)^{1/2}$, $G(f) = 10^{-10}$, y $p(t)$ = senc rt.

10.10★ (Secc. 10.1) Los filtros terminales para un sistema se han optimizado suponiendo ruido blanco y un canal sin distorsión. Sin embargo, resulta que el canal introduce alguna distorsión lineal, por lo que se agrega un igualador con $H_{eq}(f) = kH_c(f)$ después de $H_R(f)$.
(*a*) Obténgase una expresión para la $(A/\sigma)^2$ resultante en términos de $H_c(f)$ y $P(f)$ en la salida del igualador.
(*b*) Tomando a $P(f)$ como en el problema 10.8, ¿en qué factor se debe incrementar a S_T para obtener la misma probabilidad de error que si $H_c(f) = K_R$ cuando $H_c(f) = K_R/[1 + j(2f/r)]$ y se emplea $H_{eq}(f)$?.

10.11 (Secc. 10.1) Encuéntrense las ganancias de derivación c_m para un igualador de forzamiento cero de tres derivaciones cuando $p(-1/r) = 0.4$, $p(0) = 1.0$, $p(1/r) = 0.2$, y $p(k/r) = 0$ para $|k| < 1$. También calcúlese $p_{eq}(t_k)$, $-4 \leq k \leq 4$.

10.12 (Secc. 10.1) Repítase el ejercicio 10.2 con un ancho de banda disponible de 10 kHz, eligiendo parámetros para reducir al mínimo el requisito de potencia.

10.13 (Secc. 10.2) Considérese la forma de pulso *pasabanda* $p(t)$ = senc $Bt \cos 2\pi f_c t$, la cual tiene un ancho de banda B.
(*a*) Demuéstrese que $p(k/2B) = 0$ para $K \neq 0$ si $f_c = MB/2$ siendo M un entero impar. Bosquéjese a $p(t) - p(t - \frac{1}{2}B)$ tomando $M = 3$.
(*b*) Analícense las implicaciones de esto para modulación digital, incluyendo las ventajas y desventajas.

10.14 (Secc. 10.2) Intégrese la ecuación (6) sobre $0 \leq R \leq \infty$ para obtener

$$p_\phi(\phi) = \frac{A_k \cos\phi}{\sqrt{2\pi N}} e^{-(Ak2 \text{ sen } 2\phi)/2N}\left[1 - Q\left(\frac{A_k \cos\phi}{\sqrt{N}}\right)\right] + \frac{1}{2\pi} e^{-Ak2/2N}$$

y llegar a la ecuación (9) tomando a $A_b{}^2/N \gg 1$. (*Sugerencia*: Principiese con el cambio de variable $\lambda = (R - A_k \cos\phi)\sqrt{N}$ y empléese el hecho de que $Q(x) = 1 - Q(x)$.)

10.15 (Secc. 10.2) Considérese un sistema *ternario* de manipulación por corrimiento de amplitud con $A_K = 0, A$, y 24. Encuéntrese P_e en términos de ρ suponiendo que $\rho \gg$ 1. (*Sugerencia*: Justifíquese la aproximación $P_e \approx P_{e0}/3$ y nótese que $S_R = (5r/3)E$ donde E es la energía en el pulso de nivel intermedio.)

10.16 (Secc. 10.2) Supóngase que una señal binaria de manipulación por corrimiento de amplitud se detecta *en forma coherente*, por la figura 10.16, con el umbral en $A/2$. Encuéntrese P_e y compárese con las ecuaciones (15) y (21).

10.17 (Secc. 10.2) Se transmite una señal binaria de manipulación por corrimiento de frecuencia sobre un canal de radio que tiene *desvanecimiento de Rayleigh* tal que A es la variada distribuida de Rayleigh cambiando con lentitud, con $\overline{A^2} = A_0{}^2$, por lo que $p(A) = (2A/A_0{}^2)\exp(-A^2/A_0{}^2)$, $A \geq 0$.
(*a*) Demuéstrese que la probabilidad promedio de error es $E(P_e) = 1([2 + (A_0{}^2 52N)]$.
(*b*) Examínese el valor de un arreglo de diversidad como el del problema 10.3 para esta aplicación.

10.18 (Secc. 10.2) La figura P10.2 es un sistema de detección coherente para manipulación por corrimiento de fase cuaternario con $\phi_k = 0°, 90°, 180°$ y $270°$. Despreciando el ruido, constrúyase una tabla similar a la tabla 10.1, dando los valores de v_A y v_B para cada valor de ϕ_k cuando los corrimientos de fase son $\psi_A = 45°$ y $\psi_B = -45°$. Repítase para $\psi_A = 0°$ y $\psi_B = -90°$. Analícense estas dos alternativas, dando especial atención al caso donde cada valor de ϕ_k representa dos dígitos binarios.

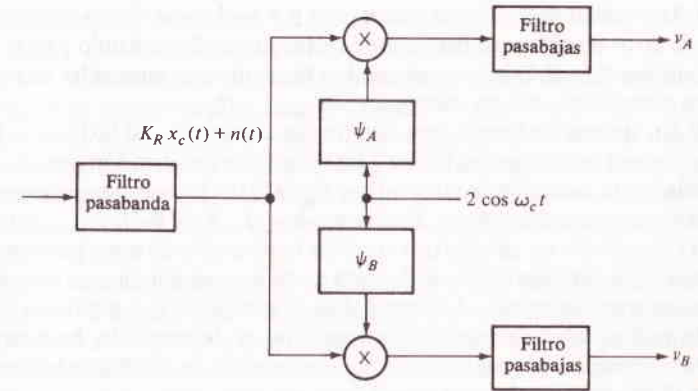



FIGURA P10.2

10.19★ (Sec. 10.2) Con relación al sistema del problema 10.18 con $\psi_A = 45°$ y $\psi_B = -45°$, escríbase el ruido $n(t)$ en la forma de la ecuación (1*a*) con $\theta = \psi_B$ y obténgase v_A y v_B cuando está presente el ruido. Demuéstrese que P(sin error) $= P(n_i \geq -A/\sqrt{2}$ y $n_q \geq -A/\sqrt{2})^e$ y de esto dedúzcase $P_e = 1 - ([1 - Q(\sqrt{p})]^2$, $p = A^2/2N$.

10.20★ (Secc. 10.2) Se va a emplear la codificación diferencial en un sistema binario polar *en banda base* para proteger contra posibles inversiones de polaridad en la transmisión. Así, si el dígito de mensaje k-ésimo es un **1**, entonces $A_k = A_{k-1}$; si el dígito k-ésimo es un **0**, entonces $A_k = -A_{k-1}$. Los valores posibles de A_k son $\pm A$. Suponiendo ruido gaussiano, demuéstrese que $P_e = [Q(A/\sigma) - Q^2(A/\sigma)]$ si la polaridad de $A_k + n_k$ se determina empleando un A/D de umbral cero y después se compara con la polaridad determinada de antemano de $A_{k-1} + n_{k-1}$ (*Sugerencia*: Ocurre un error si hay uno y sólo un error de polaridad en cada par.)

10.21★ (Secc. 10.2) Como una alternativa al esquema de detección del problema 10.20, supóngase que el receptor tiene la forma de la figura P10.3. Demuéstrese que el A/D debe tener *dos* niveles de umbral —en $\pm A$ —y por ello $P_e = 3/2\, Q(A\sqrt{2}\sigma)$. ¿Es mejor este esquema que el otro?

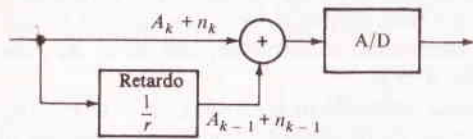
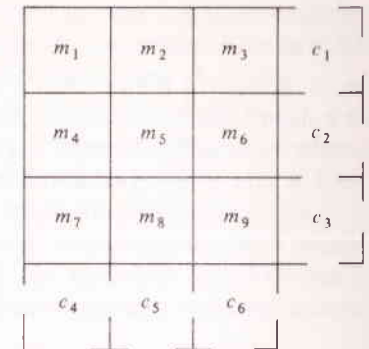



FIGURA P10.3

10.22★ (Secc. 10.2) Datos digitales con $\mu = 8$ modulan tanto la amplitud como la fase de una portadora, dando $K_R x_c(t) = A_k \cos(\omega_c t + \phi_k)$ donde $A_k = A$ y $2A$ y $\phi_k = 0°, 90°, 180°$ y $270°$.

(*a*) Dibújese un diagrama a bloques del receptor empleando un oscilador local coherente y filtros de integrado y descarga. (*Sugerencia*: Véase el problema 10.18.)

(*b*) Analícese el comportamiento de este sistema en presencia de ruido.

10.23 (Secc. 10.3) Un código determinado con $n = 7$ y $k = 3$ es en forma simultánea de corrección de error sencillo y de detección de error doble. Suponiendo $\epsilon \ll 1$, calcúlense las siguientes probabilidades: una palabra tiene un error corregido; una palabra tiene errores detectados pero no corregidos; $P_{e,\ \text{palabras}}$; $P_{e,\ \text{bit}}$.

10.24 (Secc. 10.3) Un sistema de banda base binario ideal tiene $r = 10\,000$ y $\sigma = 0.2$. Se desea enviar datos a una velocidad útil de 7 000 bits/s con codificación por comprobación de paridad para detección de error tal que $P_{\text{de}} \leq 0.08$. Encuéntrense valores apropiados para k y A, y calcúlese $P_{e,\ \text{bit}}$. *Respuesta*: $k = 3$, $A = 0.41$, $P_{e,\ \text{bit}} \leq 0.0036$.

10.25 (Secc. 10.3) Considérese un sistema que combine la *detección de error* por comprobación de paridad y la *realimentación de decisión* de tal manera que cuando se recibe una palabra con un error detectado, el receptor dice al transmisor que repita esa palabra. Si $\epsilon \leq 1$, la probabilidad de más de una repetición es despreciable. Incluyendo los efectos de la retransmisión, encuéntrese $P_{e,\ \text{bit}}$ y la relación de los dígitos totales transmitidos por dígito de mensaje en términos de k y ϵ.

10.26★ (Secc. 10.3) Un método analógico de detección de error, conocido como *detección de zona cero*, emplea una zona de no decisión centrada en cada nivel de umbral tal que si la señal más el ruido cae en esta zona, se le estima como error detectado pero no corregido.

(*a*) Obténganse expresiones para P_{de} y P_e por bit para un sistema en banda base binario polar con límites de zona cero en $\pm\, \alpha A$.

(*b*) Compárese la detección de zona cero con la detección de error por medio de la codificación por comprobación de paridad cuando ambos sistemas tienen la misma velocidad de bits de mensaje.

10.27 (Secc. 10.3) La figura P10.4 muestra el *código por comprobación de paridad en arreglo cuadrado* donde están dispuesto $k = k_0^2$ dígitos de mensaje en un cuadrado cuyas filas y columnas están comprobadas por $2\sqrt{k}$ dígitos de comprobación. Un error de transmisión en un dígito de mensaje, ocasiona que en una fila y en una columna falte la paridad con el error en el punto de intersección, por lo que el código es de corrección de error sencillo y de detección de error doble.

FIGURA P10.4

(*a*) Obténganse expresiones para $\mathscr{R}$ y $P_{e,\ \text{bit}}$ en términos de k_0.

(*b*) Analícese lo que sucede cuando hay un error en un dígito de comprobación y cuando hay dos o más errores.

10.28 (Secc. 10.3) Constrúyase una matriz **H** para un código por comprobación de paridad sistemático (6, 3) para corrección de error sencillo. Relaciónense las ecuaciones de comprobación de paridad, las palabras de código permitidas y los síndromes de error sencillo. ¿Qué sucede cuando hay errores dobles?

10.29 (Secc. 10.3) El código de Hamming del ejemplo 10.5 da indicaciones de error para los dígitos de comprobación, así como para los dígitos de mensaje, aún cuando sea innecesaria la corrección de comprobación erróneos. Sin embargo, empléese la ecuación (14) y las condiciones en **H** para comprobar que cualquier código por comprobación de paridad (n, k), para correción de error sencillo debe tener $2^{\mathcal{Q}} \geq k + 1$, por lo que los síndromes de dígitos de comprobación llegan en forma automática, ya sea que se empleen o no.

10.30 (Secc. 10.3) Un código (7, 4) de Hamming se está considerando para transmisión de datos binarios por medio de manipulación por corrimiento de frecuencia. Encuéntrese tal condición numérica que la codificación produzca una probabilidad de error más baja que la reducción de velocidad de señalización.

10.31 (Secc. 10.3) La figura P10.5 es el registro de corrimiento para un *código Hagelbarger*, uno de los primeros códigos convolucionales. El conmutador recoge en forma alternada los dígitos de mensaje y los dígitos de comprobación, por lo que $\mathscr{R} = ½$. Haciendo caso omiso del transitorio de puesta en marcha, encuéntrense los enlaces de comprobación de paridad y verifíquese que se pueden detectar seis errores sucesivos si los 19 dígitos anteriores están correctos.

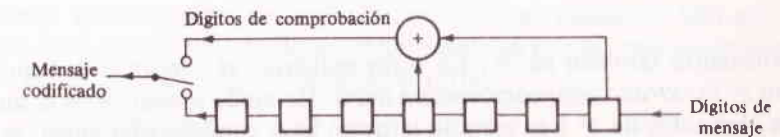



FIGURA P10.5

APÉNDICE A
EL ESPACIO DE LA SEÑAL Y LA COMUNICACIÓN

Para describir a las señales como vectores en un espacio multidimensional, se pueden aplicar las relaciones y conocimientos de geometría familiares a muchos problemas del análisis de señales y del diseño de sistemas de comunicación. Algunos resultados han sido expuestos en este texto, en forma particular la desigualdad de Schwarz y el concepto del producto escalar. Aquí se deducen esos resultados y se examinan sus aplicaciones adicionales. La presentación es breve e informal, con la menor cantidad posible de anotaciones especiales.†

A.1 TEORÍA DEL ESPACIO VECTORIAL

Un *espacio vectorial* (o espacio lineal) $\mathscr{S}$ es un conjunto de elementos llamados vectores que tienen la propiedad de que se pueden combinar en forma lineal. Así, en el caso de cualesquier vectores v y w en $\mathscr{S}$ y cualesquier escalares α y β existe otro vector

$$z = \alpha v + \beta w \tag{1}$$

el cual se encuentra también en $\mathscr{S}$. En otras palabras, el conjunto de todos los elementos en $\mathscr{S}$ es *cerrado bajo combinación lineal*. De modo similar, si $\mathscr{P}$ es un subconjunto de elementos de $\mathscr{S}$ y es cerrado también bajo combinación lineal, se dice entonces que $\mathscr{P}$ es un *subespacio* de $\mathscr{S}$.

† Ver Frederick y Carlson (1971, Caps. 5 y 12) o Wozencraft y Jacobs (1965, Caps. 4 y 5) para más detalles.

Los vectores pueden ser funciones del tiempo, incluyendo funciones complejas, mientras que los escalares son números, posiblemente complejos. A manera de ejemplo, el conjunto de todas las señales de energía $v(t)$ que tienen $\int_{-\infty}^{\infty}|v(t)|^2 dt > \infty$ constituye un espacio vectorial, puesto que es cerrado bajo combinación lineal, y el subconjunto de señales de energía restringido por $v(t) = 0$ para $t < 0$ sería un subespacio. Por otra parte, el conjunto de todas las señales limitadas por $0 \geq v(t) \leq 1$ no es cerrado bajo combinación lineal y no se califica como un espacio vectorial.

La suma vectorial y la multiplicación escalar en $\mathscr{S}$ están definidas en la manera usual y, como queda implícito por la ecuación (1), hay un vector nulo θ tal que $v + \theta = v$, $0 \cdot v = \theta$, etc. Por lo que, cada subespacio de un espacio vectorial incluye al origen.

Producto escalar y norma

Se impondrán dos condiciones adicionales en los espacios vectoriales empleados en el análisis de señales, que son para el normado y posesión de un producto escalar. El *producto escalar* es un número complejo asociado con cualquier par de vectores v y w en $\mathscr{S}$, que se simboliza por medio de $\langle v, w\rangle$, con las siguientes propiedades

$$\langle v,\theta\rangle = 0 \tag{2}$$

$$\langle v,w\rangle = \langle w,v\rangle^* \tag{3}$$

$$\langle \alpha v,\beta w\rangle = \alpha\beta^*\langle v,w\rangle \tag{4}$$

$$\langle v + w, x + y\rangle = \langle v,x\rangle + \langle v,y\rangle + \langle w,x\rangle + \langle w,y\rangle \tag{5}$$

La *norma* de cualquier vector v es un número real definido por

$$\|v\| = \langle v,v\rangle^{1/2} \tag{6}$$

con las propiedades

$$\|v\| \geq 0 \tag{7a}$$

$$\|v\| = 0 \qquad \text{sólo si} \quad v = \theta \tag{7b}$$

$$\|\alpha v\| = |\alpha|\,\|v\| \tag{8}$$

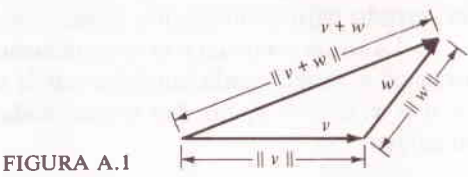
De manera formal, las condiciones anteriores caracterizan a un espacio pre-Hilbertiano. De manera intuitiva, dan una medida de la longitud del vector (la norma) y el ángulo relativo (el producto escalar). Estas interpretaciones se desarrollan en forma subsecuente. También se supone dar las fórmulas explícitas para $\langle v, w\rangle$ y $\|v\|$ puesto que dependen del espacio particular en cuestión. Aún así, la fuerza analítica de la teoría del espacio vectorial reside en que las relaciones consideradas tiene aplicaciones de cualquier espacio vectorial.

Desigualdades y ortogonalidad

Empleando las ecuaciones de la (2) a la (8) se pueden deducir varias ecuaciones importantes. Para empezar, considérese $v + \beta w$ donde v y w están en $\mathscr{S}$, β es un es-

calar arbitrario, y $w \neq \theta$. Aplicando las ecuaciones (6), (5), (4) y (3) en forma secuencial, se tiene

$$\|v + \beta w\|^2 = \langle v + \beta w, v + \beta w \rangle$$
$$= \|v\|^2 + \beta^* \langle v,w\rangle + \beta \langle v,w\rangle^* + |\beta|^2 \|w\|^2 \qquad (9)$$

FIGURA A.1

Poniendo $\beta = -\langle v, w\rangle / \|w\|^2$ y notando, de la ecuación (7a), que $\|v + \beta w\|^2 \geq 0$ para cualquier valor de β, se tiene $\|v\|^2 - (|\langle v, w\rangle| / \|w\|)^2 \geq 0$ o

$$|\langle v,w\rangle| \leq \|v\| \|w\| \qquad (10a)$$

la cual es la *desigualdad de Schwarz*. La desigualdad, obviamente, es válida en el caso trivial donde $v = \theta$ o $w = \theta$; de otra manera, a partir de la ecuación (7b) se tiene que

$$|\langle v,w\rangle| = \|v\| \|w\| \quad \text{sólo si} \quad v = \alpha w \qquad (10b)$$

donde α es arbitraria. Se puede demostrar con facilidad, a partir de las ecuaciones (9) y (10) que

$$\|v + w\| \leq \|v\| + \|w\| \qquad (11)$$

la cual se conoce como *desigualdad del triángulo*.

Por último, si $\langle v, w\rangle = 0$ por lo que $\langle v, w\rangle^* = 0$ —entonces

$$\|v + w\|^2 = \|v\|^2 + \|w\|^2 \quad \text{si} \quad \langle v,w\rangle = 0 \qquad (12)$$

Como la ecuación (12) se asemeja al teorema de Pitágoras de geometría, se dice que v y w son *ortogronales* (perpendiculares) cuando su producto escalar es igual a cero. Por lo tanto, tomando la norma como la medida de "longitud" en el espacio vectorial, las ecuaciones (11) y (12) son idénticas a las relaciones comunes para los triángulos oblicuángulos y rectángulos cuyos lados son v, w y $v + w$, (Fig. A.1). Cuando la desigualdad de Schwarz es vista bajo este punto de vista, se advierte que el producto escalar es equivalente al producto punto familiar y así se puede interpretar como una medida de "ángulo" relativo. Esta interpretación es inapreciable al considerar las proyecciones de los vectores.

Proyecciones de vectores

La figura A.2 muestra dos vectores y la *proyección ortogonal* v_w de v sobre w que se define por las propiedades $v_w = \alpha w$ y $\langle v - v_w, w\rangle = 0$, equivalente a trazar una perpendicular de v a w. Combinando estas propiedades se tiene

$$v_w = \frac{\langle v,w\rangle}{\|w\|^2} w \qquad (13)$$

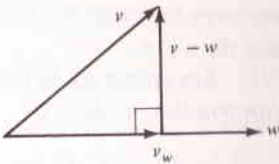



FIGURA A.2
La proyección de v sobre w.

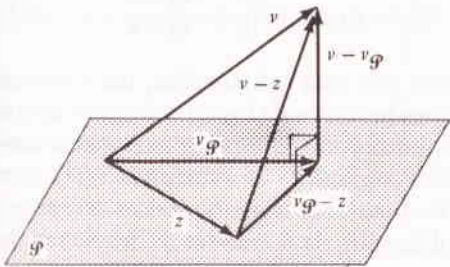



FIGURA A.3
Construcción vectorial para el teorema de la proyección.

por lo que $\langle v, w\rangle$ es proporcional al coseno del "ángulo" entre v y w; es decir, que v_w es igual a cero cuando los vectores son ortogonales y es máxima cuando son colineales (proporcionales). Se presume que $\|v - v_w\|$ es la "distancia" más corta entre el extremo de v y cualquier punto a lo largo de w.

Generalizando, si v está en $\mathscr{S}$ y $v_{\mathscr{P}}$ es su proyección en el subespacio $\mathscr{P}$, entonces $v - v$ es la "distancia" más corta entre v y cualquier punto en $\mathscr{P}$. Este es un enunciado burdo del importante y versátil *teorema de la proyección*. Establecido en forma más precisa, $v_{\mathscr{P}}$ es un vector único en $\mathscr{P}$ tal que, para cualquier vector $z \neq v_{\mathscr{P}}$ en $\mathscr{P}$,

$$\langle v - v_{\mathscr{P}}, z\rangle = 0 \qquad (14)$$
$$\|v - v_{\mathscr{P}}\| < \|v - z\| \qquad (15)$$

La ecuación (14) define a $v_{\mathscr{P}}$ mientras que la ecuación (15) es el teorema de la proyección. El teorema se prueba con facilidad empleando la construcción cuyo diagrama se aprecia en la figura A.3, donde $v - v_{\mathscr{P}}$ y $v_{\mathscr{P}} - z$ son ortogonales (por definición) y $v - z = (v - v_{\mathscr{P}}) + (v_{\mathscr{P}} - z)$. Aplicando la ecuación (12) se tiene entonces $\|v - z\|^2 = \|v - v_{\mathscr{P}}\|^2 + \|v_{\mathscr{P}} - z\|^2 > \|v - v_{\mathscr{P}}\|^2$.

Vectores base

Para cerrar esta sección, se introduce el concepto de los vectores base. Aunque no todos los espacios vectoriales tienen vectores base, se reafirman más las interpretaciones geométricas cuando ellos existen.

Supóngase un espacio que incluye un conjunto de K vectores linealmente independendientes† ϕ_k, $k = 1, 2, \ldots, K$, tal que cualquier vector v en el espacio se puede expresar en la forma

† Véase Harmuth (1969) para un análisis de las funciones de Walsh en el contexto de la comunicación.

$$v = \sum_{k=1}^{K} \alpha_k \phi_k \tag{16}$$

Se dice que el espacio tiene *K dimensiones*, simbolizado por $\mathscr{S}_K$, y que los ϕ_k son los *vectores base que cubren a* $\mathscr{S}_K$. Los escalares α_k en la ecuación (16) son las *coordenadas* de v y ϕ_k.

Sin embargo, la base para $\mathscr{S}_K$ no es única, y es ventajoso exigir que ϕ_k sean *ortonormales*, es decir,

$$\langle \phi_k, \phi_m \rangle = \begin{cases} 1 & m = k \\ 0 & m \neq k \end{cases} \tag{17}$$

lo cual asegura independencia lineal. Además, con esta estipulación, los vectores base son análogos en forma directa a los vectores unitarios de los espacios ordinarios bi y tridimensionales en los que ellos tienen norma unitaria (longitud) y son mutuamente ortogonales (perpendiculares). Los conjuntos de funciones $\phi_k(t)$ que obedecen o siguen la ecuación (17) incluyen las funciones trigonométricas que se relacionan en forma armónica, como los polinomios de Legendre y funciones de Walsh,† por nombrar unos cuantos.

La ecuación (17) no impone una limitación seria, puesto que se puede emplear el *procedimiento de Gram-Schmidt* para generar un conjunto ortonormal que cubra cualquier conjunto de vectores, como sigue. Dado el conjunto z_k, $k = 1, 2, \ldots,$ una base ortonormal equivalente es

$$\phi_k = \frac{g_k}{\|g_k\|} \tag{18a}$$

donde

$$g_1 = z_1$$

$$g_k = z_k - \sum_{m=1}^{k-1} \langle z_k, \phi_m \rangle \phi_m \qquad 2 \leq k \leq K \tag{18b}$$

Nótese, de la ecuación (13), que $\langle z_k, \phi_m \rangle \phi_m$ es en forma precisa la proyección de z_k a lo largo de ϕ_m puesto que $\|\phi_m\|^2 = 1$. Así, g_k es ortogonal a $\phi_1, \phi_2, \ldots, \phi_k - 1$ como se ilustra en la figura A.4 para $k = 2$. Si sucede que z_k no son linealmente independientes, entonces uno o más de los ϕ_k será igual a θ, significando que la dimensionalidad es menor que K.

Si se tiene una base ortonormal que cubra a $\mathscr{S}_K$, se puede expresar el producto escalar de cualesquiera dos vectores v y w en términos de sus coordenadas, digamos α_k y β_k. Introduciendo un índice m simulado,

z_2

$g_2 = z_2 - \langle z_2, \phi_1 \rangle \phi_1$

$\phi_1 = \dfrac{z_1}{\|z_1\|}$

$\langle z_2, \phi_1 \rangle \phi_1$

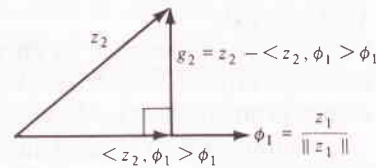



FIGURA A.4

† Véase Harmuth (1969) para un análisis de las funciones de Walsh en el contexto de la comunicación.

$$\begin{aligned} \langle v, w \rangle &= \left\langle \left(\sum_{k=1}^{K} \alpha_k \phi_k \right), \left(\sum_{m=1}^{K} \beta_m \phi_m \right) \right\rangle \\ &= \sum_{k=1}^{K} \alpha_k \left(\sum_{m=1}^{K} \beta_m^* \langle \phi_k, \phi_m \rangle \right) \\ &= \sum_{k=1}^{K} \alpha_k \beta_k^* \end{aligned} \tag{19}$$

donde se han empleado las ecuaciones (5), (4) y (17), reduciendo la última la suma sobre m a un término, β_k^*. Poniendo $\beta_k = \alpha_k$ se demuestra que $\langle v, v \rangle = \sum \alpha_k \alpha_k^*$ o

$$\|v\| = \left(\sum_{k=1}^{K} |\alpha_k|^2 \right)^{1/2} \tag{20}$$

lo cual generaliza la norma euclidiana al caso de K coordenadas valuadas complejas.

Como las coordenadas ordinarias, se les encuentra con facilidad a partir de v y ϕ_k, o sea,

$$\alpha_k = \langle v, \phi_k \rangle \qquad k = 1, 2, \ldots, K \tag{21}$$

Así, cada componente de v en la ecuación (16) es igual a la proyección de v a lo largo del correspondiente vector base, así como se pueden descomponer los vectores convencionales en sus proyecciones a lo largo de los ejes. La ecuación (21) procede de la (19) por la sustitución de w con ϕ_m por lo que $\beta_m = 1$ y $\beta_k = 0$, $k \neq m$. Por último, si v está en cualquier espacio del canal $\mathscr{P}_J$ es un subespacio J-dimensional cubierto por ϕ_k, $1 \leq k \leq J$, entonces la proyección de v sobre $\mathscr{P}_J$ es igual a

$$v_{\mathscr{P}} = \sum_{k=1}^{J} \alpha_k \phi_k \tag{22}$$

donde la ecuación (21) aún se aplica, puesto que los α_k deben ser tales que $\langle v - v_{\mathscr{P}}, \phi_k \rangle = 0$ para $k = 1, 2, \ldots, J$.

A.2 APLICACIONES DEL ESPACIO DE LA SEÑAL

Se procede ahora a las aplicaciones e interpretaciones de la teoría precedente en el análisis de señales. Esto se efectúa definiendo clases o conjuntos de señales que son cerradas bajo combinaciones lineales para cada clase, y determinando una fórmula de producto escalar apropiada. Se examinará los resultados para otras clases útiles.

Señales periódicas de potencia

Considérese el conjunto de todas las señales $v(t)$ que tienen el mismo periodo T_0 y potencia promedio bien definida. Este conjunto constituye un espacio vectorial y el producto escalar se puede tomar como

$$\langle v, w \rangle = \frac{1}{T_0} \int_{T_0} v(t) w^*(t)\, dt \tag{1}$$

Se comprueba con facilidad que la ecuación (1) tiene las propiedades necesarias, ecuaciones (2) a la (5), (Sec. A.1). La norma al cuadrado es entonces igual a la potencia promedio,

$$\|v\|^2 = \frac{1}{T_0}\int_{T0} |v(t)|^2\, dt = P_v \tag{2}$$

Así, la señal nula $\theta(t)$ es una señal periódica con $P_\theta = 0$; para todos los propósitos prácticos, esto implica que $\theta(t) = 0, -\infty < t < \infty$. De paso, también se advierte que

$$\bar{v} = \langle v,1\rangle = \frac{1}{T_0}\int_{T0} v(t)\, dt \tag{3}$$

es el valor promedio de $v(t)$.

Cualesquiera dos señales en este conjunto son colineales si son proporcionales, por ejemplo, $\|W(t)\| = \alpha v(t)$. En el otro extremo recurriendo al hecho de que $\|v\|^2 = P_v$ y a la ecuación (12), (Sec. A.1), las señales ortogonales tienen la propiedad de la superposición de la potencia promedio, es decir,

$$P_{v+w} = P_v + P_w \qquad \text{si} \quad \langle v,w\rangle = 0 \tag{4}$$

Entre estos extremos, la proyección de v a lo largo de w es

$$v_w(t) = \frac{\langle v,w\rangle}{\|w\|^2}\, w(t) \tag{5}$$

lo cual dice qué tanto de una señal, $v(t)$, está "contenida" en otra, $w(t)$.

Un subespacio importante está cubierto por la base ortonormal

$$\phi_k(t) = e^{j2\pi kt/T_0} \qquad k = 0, \pm 1, \pm 2, \ldots, \pm K \tag{6}$$

siendo la dimensionalidad $2K + 1$. Si se denota por $\tilde{v}(t)$ a la proyección de $v(t)$ en este subespacio, entonces

$$\tilde{v}(t) = \sum_{k=-K}^{K} \alpha_k \phi_k(t) = \sum_{k=-K}^{K} \alpha_k e^{j2\pi kt/T_0} \tag{7a}$$

donde, de la ecuación (21), (Sec. A.1),

$$\alpha_k = \langle v,\phi_k\rangle = \frac{1}{T_0}\int_{T0} v(t) e^{-j2\pi kt/T_0}\, dt \tag{7b}$$

La ecuación (7*a*) se reconocerá como los primeros $2k + 1$ términos de la *serie exponencial de Fourier* y α_k, las coordenadas de $\tilde{v}(t)$, son los coeficientes usuales de la serie. A partir del teorema de la proyección, se puede decir que $\tilde{v}(t)$ es una aproximación de *error cuadrático mínimo* a $v(t)$ en que la señal de error $v(t) - \tilde{v}(t)$ tiene norma mínima.

Un teorema matemático muy complejo (omitido aquí) demuestra que en el límite conforme $K \to \infty$, $\tilde{v}(t) \to v(t)$ en el sentido de que $\|v - \tilde{v}\|^2 \to 0$, conocido como la *convergencia en el medio*. En correspondencia, el subespacio resulta ser el espacio entero, con dimensionalidad infinita pero mensurable. Entonces, recurriendo a la ecuación (20), (Sec. A.1),

$$P_v = \|v\|^2 = \sum_{k=-\infty}^{\infty} |\alpha_k|^2 \tag{8}$$

la cual es el teorema de la potencia de Parseval. Se ve, pues, que los conceptos del espacio de la señal constituyen la sustentación teórica de la serie de Fourier.

Señales no periódicas de energía

El conjunto de todas las señales de energía no periódicas que tienen energía bien definida, define a un espacio de la señal. De manera similar a la ecuación (1), el producto escalar es

$$\langle v,w\rangle = \int_{-\infty}^{\infty} v(t) w^*(t)\, dt \tag{9}$$

y la ortogonalidad significa superposición de energía, puesto que $\|v\|^2 = E_v$, etc. Aunque este espacio carece de una base mensurable, se puede decir que está cubierto por

$$\phi_f(t) = e^{j2\pi ft} \qquad -\infty < f < \infty$$

donde f es una variable continua (frecuencia) más que un índice. Entonces, $v(t)$ tiene la función coordenada

$$V(f) = \langle v,\phi_f\rangle = \int_{-\infty}^{\infty} v(t) e^{-j2\pi ft}\, dt \tag{10a}$$

tal que

$$v(t) = \int_{-\infty}^{\infty} V(f) e^{j2\pi ft}\, df \tag{10b}$$

No es necesario decir que estas son las *integrales transformadas de Fourier*.

Como se recalcó en la sección 2.6, $v(t)$ y $w(t)$ son ortogonales si son ajenas (sin traslape o encimamiento) en tiempo o ajenas en frecuencia. Lo último procede de la ecuación (13), (Sec. 2.3), escrito en la presente notación como

$$\langle v,w\rangle = \langle V,W\rangle = \int_{-\infty}^{\infty} V(f) W^*(f)\, df \tag{11}$$

Así, los productos escalares de los dominios del tiempo y de la frecuencia, son iguales.

Señales de banda limitadas y señales discretas en el tiempo

Considérese el espacio definido por

$$\phi_k(t) = \sqrt{2W}\ \text{senc}\,(2Wt - k) \qquad k = 0, \pm 1, \pm 2, \ldots \tag{12}$$

así, cualquiera $v(t)$ en este espacio debe ser *de banda limitada* en W dado que

$$\Phi_k(f) = \mathscr{F}[\phi_k(t)] = \begin{cases} \dfrac{1}{\sqrt{2W}}\, e^{-j2\pi kf/2W} & |f| < W \\ 0 & |f| > W \end{cases} \tag{13}$$

es decir, las funciones base son de banda limitada así como ortonormales.

Observando que las $\phi_k(t)$ son señales de energía, la aplicación de las ecuaciones (11) y (13) da las coordenadas de $v(t)$ como

$$\alpha_k = \langle v,\phi_k\rangle = \langle V,\Phi_k\rangle = \frac{1}{\sqrt{2W}}\, v\left(\frac{k}{2W}\right) \qquad (14)$$

Por lo que,

$$v(t) = \sum_{k=-\infty}^{\infty} \alpha_k \phi_k(t) = \sum_{k=-\infty}^{\infty} v\left(\frac{k}{2W}\right) \text{senc } (2Wt - k) \qquad (15)$$

la cual desarrolla a $v(t)$ en términos de sus *valores muestra* en $t = k/2W$. El teorema del muestreo dice que la ecuación (15) es válida para cualquier señal de energía de banda limitada, como se demostró en el capítulo 8. En forma incidental, obsérvese que el espacio que contiene tales señales es de dimensionalidad infinita pero mensurable, mientras que el espacio para señales de energía de banda no limitada tiene una dimensionalidad inconmensurable.

Habiendo mencionado el teorema del muestreo, viene en forma natural a la mente el caso de las señales *discretas en tiempo*, es decir, señales que están definidas sólo en puntos discretos en el tiempo. Por lo tanto, considérese el conjunto de todas las secuencias $v(m)$, $m = 0, \pm 1, \pm 2, \ldots, \pm M$, para las cuales

$$\frac{1}{K}\sum_{m=-M}^{M} |v(m)|^2 < \infty \qquad K = 2M+1$$

Un producto escalar apropiado es

$$\langle v,w\rangle = \frac{1}{K}\sum_{m=-M}^{M} v(m)w^*(m) \qquad (16)$$

y se puede interpretar a $\|v\|^2$ como el valor cuadrático medio de la secuencia $v(m)$.

El espacio así definido tiene precisamente K dimensiones y está cubierto por

$$\phi_k(m) = e^{j2\pi km/K} \qquad k = 0, \pm 1, \ldots, \pm M$$

Las coordenadas correspondientes de $v(m)$ son

$$\alpha_k = \frac{1}{K}\sum_{m=-M}^{M} v(m)e^{-j2\pi km/K} \qquad (17)$$

la cual es una forma de la *transformada discreta de Fourier*, la teoría detrás de la transformada rápida de Fourier.

Se han considerado hasta aquí muchos de los tipos importantes de espacios de la señal —con excepción de las señales aleatorias— por lo que se cierra esta sección con dos problemas concretos que se pueden tratar y resolver con facilidad empleando los conceptos desarrollados aquí. En la siguiente sección se trata un problema más avanzado.

Síntesis de forma de onda

Supóngase que se desea un sistema capaz de generar cualquier señal de energía a partir de un conjunto dado $z_m(t)$, $m = 1, 2, \ldots, M$.

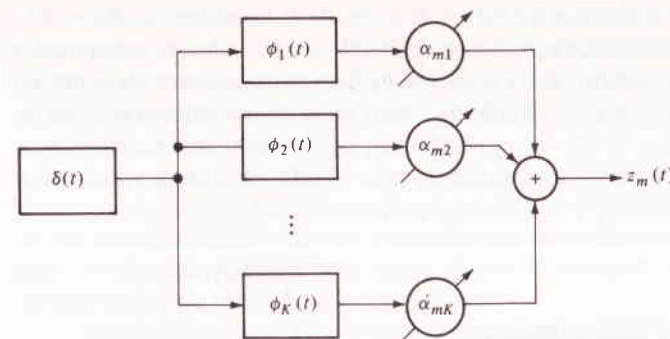



FIGURA A.5
Sintetizador de forma de onda.

Si los $z_m(t)$ no son linealmente independientes, el procedimiento de Gram-Schmidt conduce a un diseño eficiente del sistema, como sigue.

Sea $\mathscr{S}_k$ es espacio que contiene a todas las $z_m(t)$, así $K \le M$ y la igualdad es válida sólo cuando las señales son linealmente independientes. Se encuentra una base ortonormal $\phi_k(t)$ a partir de $z_m(t)$ empleando la ecuación (18), (Sec. A.1), tal que

$$z_m(t) = \sum_{k=1}^{K} \alpha_{mk}\phi_k(t) \qquad \alpha_{mk} = \langle z_m,\phi_k\rangle \qquad (18)$$

donde la fórmula para α_{mk} procede de la ecuación (21), (Sec. A.1), y el producto escalar se calcula por medio de la ecuación (9).

La implementación de la ecuación (18) incluye un generador de impulsos; K filtros que tienen a $\phi_k(t)$ como sus respuestas a impulsos, y K ganancias ajustables para dar los coeficientes α_{mk}. El sintetizador de forma de onda completo se muestra en el diagrama de la figura A.5. Si $K \ll M$, este método requiere en forma sustancial de menos componentes físicos que si se construyera un generador separado para cada una de las M señales. También es importante la ecuación (18), la cual constituye un modelo de señal útil en el estudio de la comunicación digital.

Aproximaciones lineales

Dados una señal $v(t)$ y un conjunto de K señales $w_k(t)$ sobre determinado intervalo, digamos $0 \le t \le T$, es posible construir una aproximación lineal a $v(t)$ en la forma

$$\tilde{v}(t) = \alpha_0 + \sum_{k=1}^{K} \alpha_k w_k(t) \qquad 0 \le t \le T \qquad (19)$$

con la propiedad de que $\tilde{v}(t)$ es una aproximación de error cuadrático mínimo; es decir, $\|v - \tilde{v}\|^2$ se reduce al mínimo. Si sucede que las $w_k(t)$ son ortogonales, la tarea resulta ser trivial. Si no, se podría uno inclinar por generar un conjunto ortonormal empleando el procedimiento de Gram-Schmidt. A continuación se presenta un método equivalente pero más discreto.

Primero se toma el producto escalar como

$$\langle v,w\rangle = \frac{1}{T}\int_0^T v(t)w^*(t)\,dt$$

de tal manera que el valor promedio de $v(t)$ es $\bar{v} = \langle v, 1\rangle$ y, también $\langle v, \beta\rangle = \beta^*\bar{v}$ para cualquier escalar β. Después, recordando que $v(t) - \tilde{v}(t)$ debe ser ortogonal a cada término del segundo miembro de la ecuación (19), puesto que esos términos en efecto definen el subespacio de aproximación —sean ellos ortogonales o no. Así α_0 debe satisfacer a $\langle v - \tilde{v}, \alpha_0\rangle = \langle (v - \alpha_0 - \Sigma\alpha_k w_k), \alpha_0\rangle = 0$, lo cual da $\alpha_0 = \bar{v} - \Sigma\alpha_k \bar{w}_k$. No debe sorprender que la constante α_0 es la válida para cualesquier valores promedio diferentes de cero, y la ecuación (19) resulta ser

$$\tilde{v}(t) = \bar{v} + \sum_{k=1}^{K} \alpha_k[w_k(t) - \bar{w}_k] \tag{20}$$

donde los α_k deben ser aún determinados.

Volviendo al cálculo, se emplea de nuevo la condicción de ortogonalidad, esta vez en la forma

$$\langle v - \bar{v}, \alpha_m w_m\rangle = \left\langle \left[(v - \bar{v}) - \sum_k \alpha_k(w_k - \bar{w}_k) \right], \alpha_m w_m \right\rangle$$

Una mayor simplificación da por resultado un conjunto de K ecuaciones simultáneas

$$\sum_{k=1}^{K} \langle w_k - \bar{w}_k, w_m - \bar{w}_m\rangle \alpha_k = \langle v - \bar{v}, w_m - \bar{w}_m\rangle \qquad m = 1, 2, \ldots, K \tag{21}$$

las cuales pueden ser resueltas para α_k. En el caso especial de $K = 1$, la aproximación completa es

$$\tilde{v}(t) = \bar{v} + \frac{\langle v - \bar{v}, w - \bar{w}\rangle}{\|w - \bar{w}\|^2}[w(t) - \bar{w}] \tag{22}$$

El método anterior también es aplicable a señales aleatorias o a datos experimentales, tomándose el producto escalar como $\langle v, w\rangle = E[v, w^*]$. En ese contexto se le conoce como análisis *de regresión lineal.*

A.3 TEORÍA DE LA DETECCIÓN

Shannon (1949) exploró los conceptos del espacio de la señal para encontrar la capacidad de un canal continuo. Después, Arthurs y Dym (1962) y otros emplearon interpretaciones geométricas similares para estudiar la detección de señales digitales contaminadas con ruido, obteniendo de ahí la implementación de un sistema cuyo comportamiento se aproxima al límite de Shannon. Esta sección sirve como introducción a una pequeña, pero importante, división de la teoría de la detección basada en el trabajo original de ellos.

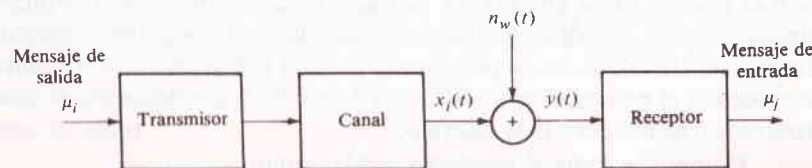



FIGURA A.6

Con referencia a la figura A.6, de una fuente de información se selecciona un mensaje μ_i de un conjunto de M mensajes equiprobables. El mensaje se transmite como una forma de onda real, produciendo $x_i(t)$ en el receptor; $x_i(t)$ tiene una duración T, digamos $0 \le t \le T$, y energía E_i. Se contamina con ruido blanco aditivo gaussiano de valor medio cero $n_w(t)$ con $G_{n_w}(f) = n/2$. El receptor opera en $y(t) = x_i(t) + D_w(t)$ y determina que el mensaje μ_j ha sido enviado, el cual, puede o no ser correcto. Se supone que el receptor tiene copias almacenadas utilizables de las señales no contaminadas y de sus mensajes correspondientes. El problema de la detección resulta ser este: ¿Cómo debe operar el receptor sobre $y(t)$ para reducir la probabilidad de error P_e al mínimo?

Obsérvese que se han supuesto mensajes equiprobables y ruido blanco gaussiano. Una codificación de fuente apropiada puede asegurar lo primero y en forma eventual se permitirá para ruido no blanco gaussiano, por lo que el enunciado del problemas es razonablemente amplio. Además, no hay preconcepciones incorporadas acerca de la estructura del receptor junto a la suposición de las copias almacenadas. La formulación del espacio de la señal del problema, será en esencia lo que guíe la implementación específica del receptor.

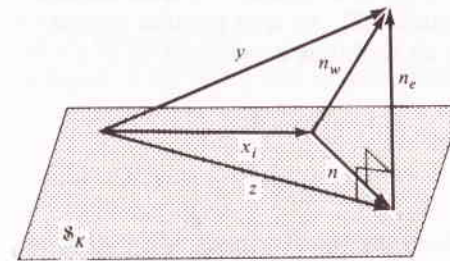



FIGURA A.7
Proyección de $y = x_i + n_w$ sobre $\mathcal{S}_k$.

Formulación del espacio de la señal

Recurriendo al procedimiento de Gram-Schmidt, existe un espacio $\mathcal{S}_k$ de dimensionalidad $K \le M$ que contiene todas las formas de onda de señal posibles, $x_i(t)$, $i = 1, 2, \ldots, M$. Supóngase que $\mathcal{S}_k$ está cubierta por una base ortonormal tal que

$$x_i(t) = \sum_{k=1}^{K} x_{ik}\phi_k(t) \tag{1a}$$

donde

$$x_{ik} = \langle x_i, \phi_k\rangle = \int_{-\infty}^{\infty} x_i(t)\phi_k(t)\,dt \tag{1b}$$

empleando el producto escalar de energía con $\phi_k(t)$ real, en vista de que todas las $k_i(t)$ son reales. (En consecuencia, los términos complejos conjugados no aparecerán en alguna parte.)

Como el ruido es blanco y de banda no limitada, sus funciones muestra $n_w(t)$ no están contenidas en $\mathcal{S}_k$. Sin embargo, se le puede descomponer en dos términos

$$n_w(t) = n(t) + n_e(t) \tag{2}$$

de los cuales $n_e(t)$ está fuera de $\mathcal{S}_k$, es decir,

$$\langle n_e, \phi_k \rangle = 0 \qquad k = 1, 2, \ldots, K \tag{3}$$

Por lo que a $n_e(t)$ se le puede designar como ruido extraño o *irrelevante*. El término restante $n(t)$ está en $\mathcal{S}_k$ y

$$n(t) = \sum_{k=1}^{K} n_k \phi_k(t) \tag{4a}$$

$$n_k = \langle n, \phi_k \rangle = \langle n_w, \phi_k \rangle \tag{4b}$$

puesto que, de las ecuaciones (2) y (3), $\langle n, \phi_k \rangle = \langle n_w, \phi_k \rangle - \langle n_e, \phi_k \rangle = \langle n_w, \phi_k \rangle$.

Estas manipulaciones conducen al diagrama vectorial de la figura A.7 donde, por conveniencia, $\mathcal{S}_k$ está representado por un plano. El vector señal x_i está en el plano como también el vector ruido n, en tanto que n_e es ortogonal a aquél Se define un nuevo vector

$$z \triangleq y - n_e = x_i + n \tag{5}$$

el cual viene a ser la *proyección* de $y = x_i + n_w$ sobre el espacio $\mathcal{S}_k$. Esta definición es valiosa de manera particular, con la ecuación (2), ya que permite analizar el problema de la detección en forma completa en términos de vectores en $\mathcal{S}_k$.

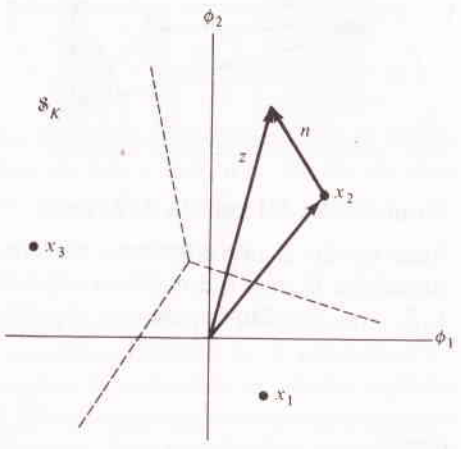



FIGURA A.8
Regiones de decisión en $\mathcal{S}_k$; $k = 2, M = 3$.

La figura A.8 ilustra este último punto al mostrar a $\mathcal{S}_k$ para el caso de $K = 2$ y $M = 3$. Los tres puntos representan los extremos de los vectores de señal libres de ruido x_1, x_2, y x_3, y z es una proyección vectoria típica cuando se envía a x_2 contaminada por el ruido n. En forma intuitiva, el diagrama sugiere que el receptor debe elergir el vector de señal cuyo extremo está más cerca del extremo de z, es decir, elegir (o detectar) a μ_j si

$$\|z - x_j\| < \|z - x_i\| \quad \text{toda} \quad i \neq j \tag{6}$$

Se demostrará además que ésta es la estrategia óptima de detección bajo las suposiciones enunciadas. En correspondencia con la ecuación (6), se puede dividir a $\mathcal{S}_k$ en *M regiones de decisión* limitadas por las líneas punteadas en la figura A.8.

La regla de decisión (6) incluye longitudes de vector o normas. Se sabe que $\|x_i\|^2 = E_i$, pero $n(t)$ es una variable aleatoria cuyas propiedades necesitan ser encontradas antes de seguir adelante.

De la ecuación (4*a*) la norma al cuadrado de $n(t)$ es

$$\|n\|^2 = \sum_{k=1}^{K} n_k^2 \tag{7}$$

y, puesto que la ecuación (4*b*) es una operación lineal en una variada gaussiana, las coordenadas n_k son también gaussianas con $\bar{n}_k = \langle \bar{n}_w, \phi_K \rangle = 0$. Se vuelve entonces a la variancia y a la cuestión de la dependencia estadística al considerar

$$\begin{aligned} \mathrm{E}[n_k n_m] &= \mathrm{E}\left[\int_{-\infty}^{\infty} n_w(t)\phi_k(t)\, dt \int_{-\infty}^{\infty} n_w(\lambda)\phi_m(\lambda)\, d\lambda\right] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{E}[n_w(t)n_w(\lambda)]\phi_k(t)\phi_m(\lambda)\, dt\, d\lambda \end{aligned}$$

Ahora $\mathrm{E}[n_w(t)n_w(\lambda)] = R_{nw}(t - \lambda)$ y la función de autocorrelación del ruido blanco es $R_{n_w}(\tau) = (\eta/2)\delta(\tau)$ por lo que, después de algunas manipulaciones sencillas,

$$\mathrm{E}[n_k n_m] = \begin{cases} \dfrac{\eta}{2} & m = k \\ 0 & m \neq k \end{cases} \tag{8}$$

por lo que las n_k son variadas gaussianas *estadísticamente independientes* con valor medio cero y variancia $\eta/2$. En concordancia, la función de densidad de probabilidad conjunta del vector de ruido n es

$$\begin{aligned} p_n(n) &= p_{n_1}(n_1)p_{n_2}(n_2) \cdots p_{n_K}(n_K) \\ &= (\pi\eta)^{-K/2} e^{-\|n\|^2/\eta} \end{aligned} \tag{9}$$

la cual tiene simetría esférica en $\mathcal{S}_k$.

Regla de decisión

Para apoyar en forma analítica la regla de decisión intuitiva, se advierte que la probabilidad de error de detección se reduce al mínimo si μ_j es el mensaje intentado más probable dado el vector recibido z, es decir,

$$P(\mu_j | z) > P(\mu_i | z) \quad \text{toda} \quad i \neq j \tag{10}$$

Tales probabilidades condicionales se pueden escribir empleando la regla mixta de Bayes en la forma

$$P(\mu_i | z) = P(\mu_i) \frac{p_z(z | \mu_i)}{p_z(z)} \tag{11}$$

donde $p_z(z\ \mu_i)$ es la función de densidad de probabilidad condicional de z dado que x_i

fue enviada, etc., y $p(\mu_i) = 1/M$ por suposición anterior. Puesto que $z = x_i + n$, $n = z - x_i$ y

$$p_z(z|\mu_i) = p_n(z - x_i) = (\pi\eta)^{-K/2} e^{-\|z - x_i\|^2/\eta} \qquad (12)$$

Por lo tanto, insertando las ecuaciones (12) y (11) en la (10) y tomando el logaritmo de ambos miembros se tiene

$$\|z - x_j\|^2 < \|z - x_i\|^2 \quad \text{toda } i \neq j \qquad (13)$$

todos los otros términos se cancelan. La ecuación (13) es equivalente a la (6) dado que las normas son no negativas por definición.

Por desgracia, no es posible diseñar un receptor en forma directa a partir de la ecuación (13); el receptor debe operar en $y(t) = x_i(t) + n_w(t)$ y no en la proyección vectorial z. Sin embargo,

$$\|z - x_i\|^2 = E_z - 2\langle z, x_i\rangle + E_i$$

y

$$\langle z, x_i\rangle = \langle y - n_e, x_i\rangle = \langle y, x_i\rangle$$

Sustituyendo éstas en la ecuación (13) y despreciando cualesquier términos independientes de i se tiene la útil prueba estadística o *función de decisión*

$$\mathscr{D}_i = \langle y, x_i\rangle - \tfrac{1}{2}E_i \qquad (14)$$

y la regla de decisión resulta ser: Elíjase a μ_j tal que $\mathscr{D}_j < \mathscr{D}_i$ para toda $i \neq j$. Ésta es una regla de decisión óptima en el sentido de que reduce al mínimo a P_e.

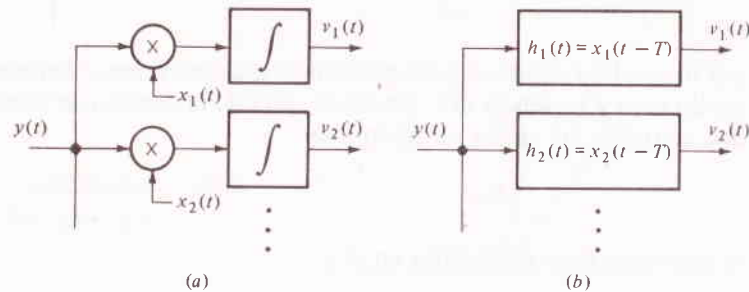



FIGURA A.9
Implementación de receptor empleando: (*a*) correlacionadores; (*b*) filtros acoplados.

La implementación del receptor

Se implementa la ecuación (14) observando que $x_i(t) = 0$ para $t < 0$ o $t > T$ por lo cual

$$\langle y, x_i\rangle = \int_0^T y(t)x_i(t)\,dt \qquad (15)$$

Por lo tanto, y en atención a la figura A.9*a*, el receptor debe multiplicar a $y(t)$ por las copias almacenadas de cada posible $x_i(t)$ e integrar, produciendo M salidas en paralelo $v_i(T) = \langle y, x_i\rangle$. Este proceso se conoce como *detección de correlación*. De manera alterna, el banco de correlatores se puede reemplazar por un banco de *filtros acoplados*, siendo las respuestas al impulso $h_i(t) = x_i(T - t)$, (Fig. A.9*b*). Entonces

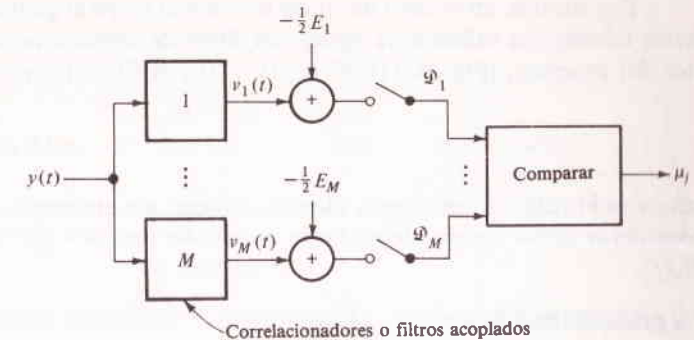



FIGURA A.10
Receptor completo para detección óptima.

$$v_i(t) = [y(t)] * [h_i(t)] = \int_0^T y(\lambda)x_i(T - t + \lambda)\,d\lambda \qquad (16)$$

y de nuevo $v_i(T) = \langle y, x_i\rangle$. De cualquier manera, la operación $\langle y, x_i\rangle$ proyecta a lo largo de x_i y remueve el ruido irrelevante $n_e(t)$.

La figura A.10 muestra un receptor completo que utiliza correlatores o filtros acoplados. La señal entrante se procesa en paralelo por medio del banco correlacionador/filtro cuyas salidas están polarizadas por $-\tfrac{1}{2}E_i$. (La polarización se puede omitir si todas las señales tienen la misma energía.) Muestreando en $t = T$ se tiene $v_i(T) - \tfrac{1}{2}E_i = \mathscr{D}_i$ y el receptor elige el mensaje en base al mayor $\mathscr{D}_i$.

Si las señales $x_i(t)$ no son linealmente independientes y $K \ll M$, un diseño más eficiente se basa en la escritura de

$$\langle y, x_i\rangle = \sum_{k=1}^{K} y_k x_{ik} \qquad y_k = \langle y, \phi_k\rangle \qquad (17)$$

Así, y_k se genera a partir de $y(t)$ y de $\phi_k(t)$ por medio de la correlación y filtraje de acoplamiento con K en lugar de M unidades. El receptor calcula entonces

$$\mathscr{D}_i = \sum_{k=1}^{K} y_k x_{ik} - \tfrac{1}{2}E_i$$

y procede como anteriormente. De manera incidental, sólo para esta implementación se necesita conocer los $\phi_k(t)$ en forma explícita.

$G_n(f)$ — $|H_K(f)|^2 = \dfrac{1}{G_n(f)}$ — $x_i(t) + n_w(t)$

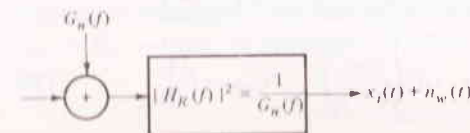



FIGURA A.11
Filtro de preblanqueado.

Por último, en el caso de ruido no blanco (pero sí gaussiano), el resultado anterior también es válido si se agrega un *filtro de preblanqueado* en el extremo anterior del receptor, (Fig. A.11). Si $G_n(f) \neq 0$ y el filtro tiene

$$|H_R(f)|^2 = \frac{1}{G_n(f)}$$

entonces el ruido filtrado será blanco. Nótese, sin embargo, que a $x_i(t)$ se les debe considerar como formas de onda de las señales después que han pasado a través de $H_R(f)$.

La probabilidad de error

Para concluir este tratado abreviado de la teoría de la detección, se examina la probabilidad de error empleando siempre argumentos geométricos. En particular, la figura A.12 representa la situación en $\mathscr{S}_k$ cuando se ha enviado a x_j y contaminado con ruido n para producir el vector z. También se muestra otro vector de señal x_i y la proyección de n a lo largo del vector $x_i - x_j$. La longitud de esa proyección (incluyendo el signo algebraico) es

$$n_{ij} = \frac{\langle n, x_i - x_j \rangle}{L_{ij}} \tag{18a}$$

donde

$$L_{ij} = \|x_i - x_j\| \tag{18b}$$

Es obvio que resulta un error de detección si n ocasiona que z esté más cerca de x_i que de x_j, o en forma equivalente,

$$n_{ij} \geq \tfrac{1}{2} L_{ij} \qquad i \neq j$$

Se concluye, a partir del análisis que condujo a la ecuación (8) que la proyección de n a lo largo de *cualquier* vector en $\mathscr{S}_k$ es una variada gaussiana con valor medio cero y variancia $\eta/2$, por lo que

$$\text{Prob}\,[n_{ij} \geq \tfrac{1}{2} L_{ij}] = Q\left(\frac{L_{ij}}{\sqrt{2\eta}}\right)$$

En consecuencia, computando para todos los $M-1$ vectores de señal x_i, $i \neq j$, la probabilidad de error cuando x_j se envía tiene un límite superior dado por

$$P_{e_j} \leq (M-1) Q\left(\frac{L_j}{\sqrt{2\eta}}\right) \qquad L_j = L_{ij}\Big|_{\substack{\min \\ i \neq j}} \tag{19}$$

es decir, que L_j es la distancia entre x_j y su vecino más cercano.

Promediando la ecuación (19) con todas las señales posibles se tiene

$$P_e = \frac{1}{M} \sum_{j=1}^{M} P_{e_j} \leq \frac{M-1}{M} \sum_{j=1}^{M} Q\left(\frac{L_j}{\sqrt{2\eta}}\right)$$

$$\leq (M-1) Q\left(\frac{L_{\min}}{\sqrt{2\eta}}\right) \qquad L_{\min} = L_j\Big|_{\min} \tag{20}$$

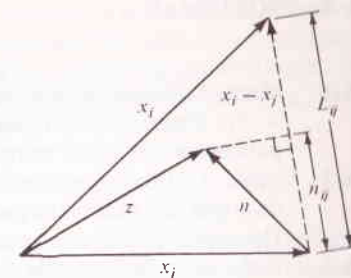



FIGURA A.12

el cual es un límite superior conservativo, puesto que $L_{\min}$ es el valor más pequeño de L_j. Arthurs y Dym después van a demostrar que P_e tiene un límite inferior dado por

$$P_e \geq Q\left(\frac{L_{\text{av}}}{\sqrt{2\eta}}\right) \qquad L_{\text{av}} = \frac{1}{M} \sum_{j=1}^{M} L_j \tag{21}$$

La figura A.13 muestra estas diferentes distancias tomando x_1, x_2, y x_3 de la figura A.8. Para un sistema binario ($M = 2$),

$$P_e = Q\left(\frac{\|x_1 - x_2\|}{\sqrt{2\eta}}\right) \tag{22}$$

puesto que $L_{\text{prom}} = L_{\min} = \|x_1 - x_2\|$.

Además de dar los límites de probabilidad de error con detección óptima, las ecuaciones (20) y (21) constituyen unas valiosas guías para la *selección de la señal.* † Como ejemplo, considérese un sistema binario con una energía limitada por $E_i \leq E$; los diagramas vectoriales sencillos comprueban que $\|x_1 - x_2\|$ se aumenta al máximo (con lo que P_e se reduce al mínimo) si $x_2(t) = -x_1(t)$ y $E_1 = E_2 = E$. Además, se

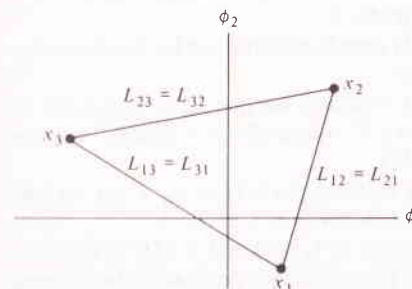



FIGURA A.13
Distancias en $\mathscr{S}_k$. Puesto que $L_{12} < L_{13} < L_{23}$, $L_1 = L_2 = L_{12}$, $L_3 = L_{13}$, L_1, y L_{prom}, $L_{j\min} = (2L_{12} + L_{13})/3$.

puede demostrar que ‡ el de M señales *ortogonales* con igual energía E y con la implementación de receptor analizada con anterioridad, se llega a un sistema cuyo comportamiento se aproxima al de un sistema *ideal* (de acuerdo con Shannon) que tiene un ancho de banda infinito; es decir, que conforme $M \to \infty$, $P_e \to 0$ siempre y cuando la velocidad de información satisfaga a $\mathscr{R} = (1/T)\log_2 M \leq 1.44_R/\eta$, donde $S_R = ET$ —véase la ecuación (7), (Sec. 9.4).

† Véase Lathi (1968, Cap. 6) o Wozencraft y Jacobs (1965, Cap. 5).
‡ Lathi (1968, Cap. 7) o Wozencraft y Jacobs (1965, Cap. 5).

A.4 PROBLEMAS

A.1 (Secc. A.1) Considérense los vectores v, w, y $z = v + w$ con $\|v\|^2 = \|w\|^2 = 16$ y $\|z\|^2 = 32$. Constrúyase un diagrama vectorial basado en esta información y empléesele para encontrar a $\|x\|^2$ cuando $x = 3\ v = 2_w$. (*Respuesta*: $\|x\|^2 = 208$.)

A.2 (Secc. A.1) En referencia al problema A.1, encuéntrese un vector y en términos de v y de w que esté en el plano que contiene a v y a w y tiene la propiedad de que $\langle y, z \rangle = 0$. Después encuéntrense las proyecciónes y_v y y_w.

A.3 (Secc. A.1) Dedúzcase la desigualdad del triángulo en la forma $\|v + w\|^2 \leq (\|v\| + \|w\|)^2$ empezando con la ecuación (9) y aplicando la (10). (*Sugerencia*: Nótese que $\langle v, w \rangle + \langle v, w \rangle^* = 2\ \mathrm{Re}[\langle v, w \rangle]$.)

A.4★ (Secc. A.1) Haciendo $v = \alpha w + \beta z$, donde z es un vector arbitrario, encuéntrese la relación entre v y w tal que $\|v + w\| = \|v\| + \|w\|$. ¿Está el resultado de acuerdo con la gráfica vectorial?

A.5 (Secc. A.1) Supóngase que $v = \alpha_1 z_1 + \alpha_2 z_2$ y $w = \beta_1 z_1 + \beta_2 z_2$ donde z_1, z_2 no son ortonormales. Encuéntrese una expresión para $\langle v, w \rangle$ y compáresele con la ecuación (19).

A.6 (Secc. A.2) Considérese el espacio cubierto por las funciones periódicas $\phi_1(t)$ y $\phi_2(t)$ donde $\phi_1(t)$ es la onda cuadrada de la figura 2.10 con $A = 1$ y $\phi_2(t) = \phi_1(t - T_0/4)$.
(*a*) Demuéstrese que $\phi_1(t)$ y $\phi_2(t)$ son ortonormales.
(*b*) Encuéntrense α_1 y α_2 tales que $\tilde{v}(t) = \alpha_1\phi_1(t) + \alpha_2\phi_2(t)$ es la aproximación de error cuadrático mínimo a $v(t) = A \cos(2\ \pi t/T_0 + 45°)$.

A.7 (Secc. A.2) Si $v(t)$ es una señal periódica real, se le puede desarrollar en la forma $v(t) = a_0 + \Sigma^{\infty} = 1\ (a_n \cos 2\pi nt/T_0 + b_n \operatorname{sen} 2\pi nt/T_0)$, la *serie de Fourier del seno y del coseno*. Identifíquense las funciones bases, compruébense para ortonormalidad, y obténganse fórmulas para a_0, a_n y b_n en términos de $v(t)$.

A.8 (Secc. A.2) Investíguense las propiedades de la base de señal de energía $\phi_f(t) = e^{j2\pi ft}$ considerando $\langle \phi_{f_1}, \phi_{f_2} \rangle = \lim_{T\ \infty} \int_{-T}^{T} \phi_{f_1}(t)\phi_{f_2}^*(t)dt$.

A.9 (Secc. A.2) Empléense las ecuaciones (11) y (13) para demostrar que las funciones bases dadas en la ecuación (12) son ortonormales.

A.10★ (Secc. A.2) Tomando α_k por la ecuación (17), demuéstrese que $\Sigma_{k = -M}^{M}\alpha_k\phi_k(n) = v(n)$. (*Sugerencia*: Nótese que $\exp j2\pi k(n - m)/K = [\exp j2\pi(n - m)/K]^k$ e introdúzcase un nuevo índice de suma $k' = k + K$.)

A.11 (Secc. A.2) Obtenganse las fórmulas para α_0 y α_1 tales que $\tilde{v}(t) = \alpha_0 + \alpha_1 t$ sea una aproximación de error cuadrático mínimo a $v(t)$ sobre $0 \leq t \leq T$.

A.12★ (Secc. A.2) Considérense las señales de energía $z_1(t) = 1$, $z_2(t) = t$, y $z_3(t) = t^2$ definidas sobre el intervalo $-1 \leq t \leq 1$. Empléese el procedimiento de Gram-Schmidt para generar las correspondientes funciones base ortonormales. Los resultados serán proporcionales a los tres primeros *polinomios de Legendre*. (*Respuesta*: $\phi_1(t) = 1/\sqrt{2}$, $\phi_2(t) = \sqrt{3/2}\,t$, $\phi_3(t) = \sqrt{5/8}(3t^2 - 1)$.)

A.13★ (Secc. A.2) Dada una señal de energía $v(t)$, se desea construir la *aproximación constante en grado mínimo* $\tilde{v}(t) = \Sigma^{\infty}{}_{k = -\infty} a_k \Pi[(t - k\tau)/\tau]$. Obténgase una fórmula para a_k en términos de $v(t)$ de tal manera que $\tilde{v}(t)$ sea una aproximación de error cuadrático mínimo.

A.14 (Secc. A.3) Obténgase una versión más general de la ecuación (14) partiendo de la (10) y prescindiendo de la suposición de mensajes equiprobables, por lo que los $P(\mu_i)$ son arbitrarios. (*Respuesta*: $\mathcal{D}_i = \langle y, x_i \rangle + [\eta \ln p(\mu_i) - E_i]/2$.)

A.15 (Secc. A.3) Encuéntrese P_e en términos de $p = E/\eta$, donde E es la *energía promedio*, para cada caso relaciónese α con E, suponiendo que α es real y que $\phi_1(t)$ y $\phi_2(t)$ son una ortonormal pero la otra arbitraria, y ubíquese a x_1, x_2, y la frontera de la región de decisión en el plano $\phi_1 - \phi_2$; ¿compárense los resultados con la ecuación (17*b*), (Sec. 4.5).
(*a*) $x_1(t) = \alpha\phi_1(t), x_2(t) = 0$;
(*b*) $x_1(t) = \alpha\phi_1(t), x_2(t) = \alpha\phi_2(t)$;
(*c*) $x_1(t) = \alpha\phi_1(t), x_2(t) = -\alpha\phi_1(t)$.

A.16 (Secc. A.3) Considérese un sistema con $M = 5$ en el cual $x_1(t) = 0$, $x_2(t) = \phi_1(t)$, $x_3(t) = \phi_2(t)$, $x_4(t) = -\phi_2(t)$, y $x_5(t) = -\phi_2(t)$, donde ϕ_1 y ϕ_2 son ortonormales. Trácense los vectores y las fronteras de la región de decisión en el plano $\phi_1 - \phi_2$, y encuéntrense L_j, $L_{\min}$ y L_{prom}.

A.17★ (Secc. A.3) Dado un conjunto de M vectores de señal equiprobables en $\mathcal{S}_K$, la energía promedio es $E = (1/M)\Sigma_{i=1}^{M} \|x_i\|^2$. Supóngase que se construye un nuevo conjunto al sustraer un vector fijo $+ = \Sigma_{k=1}^{k} \alpha_k\phi_k$, es decir, $x_i' - w$. Esto no cambia la probabilidad de error (¿por qué?), pero la energía promedio es ahora $E' = (1/M)\Sigma_{i=1}^{M} \|x_i - w\|^2$. Suponiendo que todos los vectores y escalares son reales, demuéstrese que a E'' se le reduce al mínimo si las coordenadas de w son $\alpha_m = (1/M)\Sigma_{i=1}^{M} x_{im}$, $m = 1, 2, \ldots, K$. Al conjunto x_i' se le designa entonces como el *conjunto equivalente de energía mínima*. (*Sugerencia*: Desarróllese a $\|x_i - w\|^2$ y tómese la derivada parcial de E'' con respecto a α_m.)

APÉNDICE B
RUIDO EN EL RECEPTOR

El estudio del ruido en la comunicación eléctrica es un tema vasto y multifacético, porque el ruido procede de una variedad de mecanismos y entra al sistema en cualquier punto. Sin embargo, los efectos son más serios donde el nivel de la señal es más bajo en el receptor. En consecuencia, por la figura B.1, el problema se puede dividir en tres partes: (1) la relación señal a ruido de la entrada o fuente $(S/N)_S$ en las terminales de la antena (las terminales de una línea de transmisión en el caso de sistema alámbricos); (2) la relación señal a ruido de la predetección $(S/N)_R$; y (3) la relación señal a ruido en el destino $(S/N)_D$.

Para empezar con la entrada, el ruido entrante es por lo general el resultado de la radiación electromagnética interceptada por la antena, siendo esta radiación tanto de procedencia humana, como de fuentes naturales. Ejemplos del primer tipo son las chispas de la ignición o de los conmutadores, lámparas fluorescentes, máquinas de rayos X; del segundo son el ruido atmosférico (estática) y emisiones de cuerpos extraterrestres.† Por añadidura, la antena por sí misma puede ser un generador de ruido. Como resultado, la potencia de ruido entrante depende del tipo de antena, su lugar de ubicación y orientación, frecuencia y horario de operación, y así por el estilo. Es obvio, que no es posible hacer aquí una investigación de todos estos factores; además, aunque existen fórmulas de predicción, la determinación exacta, precisa, del ruido de entrada requiere por lo general que las mediciones sean hechas *in situ*. Se supondrá, por lo tanto, que la relación señal a ruido de entrada $(S/N)_S$ es conocida, por un medio u otro.

† Tales emisiones, conocidas como *ruido cósmico de radio*, fueron descubiertas por Karl Jansky en 1932, y así nació el campo de la radioastronomía. Se ha aprendido mucho acerca de nuestro universo estudiando este ruido. No obstante, constituye una molestia para el ingeniero de comunicaciones.

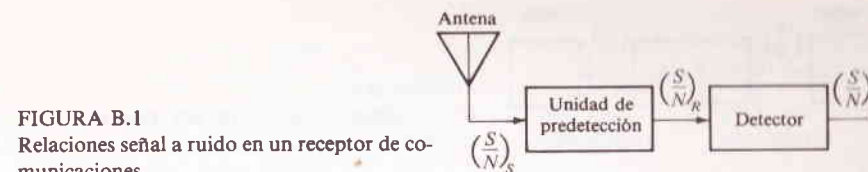



FIGURA B.1
Relaciones señal a ruido en un receptor de comunicaciones.

En el otro extremo del receptor, el detector y demás etapas subsecuentes están casi siempre libres de ruido, por lo que la relación entre $(S/N)_R$ y $(S/N)_D$ depende sólo del tipo del sistema de modulación. Estas relaciones se desarrollaron en los capítulos 7 y 8.

Si la unidad de predetección está también exenta de ruido, entonces $(S/N)_R$ es igual a $(S/N)_S$ y el sistema no presenta problema. Pero este caso es muy raro, puesto que las etapas de RF y FI, la del mezclador, etc., a menudo agregan ruido que se genera internamente y cuyo nivel de potencia es comparable, o incluso mayor que el ruido de entrada. Para una evaluación precisa del comportamiento del sistema, se debe entonces incluir el *ruido del receptor*. Y esa es la tarea a ejecutar.

En este Apéndice se analiza el ruido del receptor en términos de las mediciones típicas de la *figura y temperatura del ruido*, mediciones que facilitan el diseño del sistema y la evaluación al relacionar a $(S/N)_R$ con $(S/N)_S$. Recurriendo a los conceptos del ruido térmico, potencia disponible y ancho de banda equivalente de ruido de la sección 3.6 (la cual es conveniente se revise por parte del lector antes de proseguir), se desarrollará la descripción de las redes ruidosas de dos puertas. Después se demostrará que las primeras etapas de un receptor son los elementos críticos en el diseño de bajo ruido. Sin embargo, el análisis del ruido en los dispositivos en sí, queda fuera del alcance de este texto.†

B.1 FIGURA DE RUIDO Y TEMPERATURA EFECTIVA DE RUIDO

La porción de predetección de un receptor consta de varias unidades conectadas en cascada, en cada una de las cuales se tiene ruido generado en forma interna. Esta sección trata en forma completa de la descripción de una red de dos puertas, ruidosa, sencilla (Fig. B.2). Sea $G_{si}(f) + G_{ni}(f)$ la densidad espectral de una señal más ruido en la entrada y sea $H(f)$ la función de transferencia. Si $G_{nx}(f)$ es la densidad de salida del *ruido excedente* introducido por el propio transductor de dos puertas, entonces la relación señal a ruido de salida es

$$\left(\frac{S}{N}\right)_o = \frac{\int_{-\infty}^{\infty} |H(f)|^2 G_{si}(f)\, df}{\int_{-\infty}^{\infty} [|H(f)|^2 G_{ni}(f) + G_{nx}(f)]\, df} \qquad (1)$$

$G_{si}(f) + G_{ni}(f)$ → $H(f)$, $G_{nx}(f)$ → $|H(f)|^2\, [G_{si}(f) + G_{ni}(f)] + G_{nx}(f)$

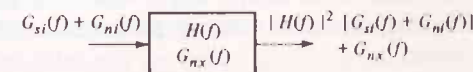



FIGURA B.2

† Véase para referencia, la lectura suplementaria seleccionada.

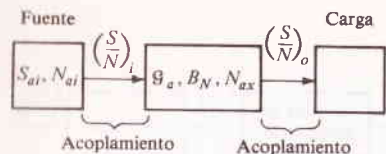



FIGURA B.3
Un transductor ruidoso de dos puertas con fuente y carga acopladas.

Esta expresión formidable se simplifica mucho si el ruido de entrada es blanco, o si, por lo menos, tiene densidad uniforme η sobre la pasabanda del dispositivo, y la relación de amplitudes es constante de manera esencial, por lo que $|H(f)|^2 = \mathcal{G}$ sobre el intervalo de frecuencia de la señal de entrada. Bajo estas condiciones

$$\left(\frac{S}{N}\right)_o = \frac{\mathcal{G}\int_{-\infty}^{\infty} G_{si}(f)\,df}{\eta\int_0^{\infty}|H(f)|^2\,df + \int_{-\infty}^{\infty} G_{nx}(f)\,df} = \frac{\mathcal{G}S_i}{\mathcal{G}\eta B_N + N_x} \tag{2}$$

donde S_i es la potencia de la señal de entrada y Nx es la potencia total de ruido excedente en la salida. La ecuación (2) hace resaltar el hecho de que un transductor de dos puertas amplifica (o atenúa) la señal de entrada y al ruido en la misma cantidad y además agrega ruido por sí mismo.

Temperatura efectiva de ruido

No obstante las simplificaciones, la ecuación (2) es cuando mucho una descripción engorrosa de una fuente de ruido de dos puertas. Si por el momento se supone que todas las impedancias están acopladas, se obtiene una forma de medir más manejable; es decir, que viendo desde dentro del transductor de dos puertas, la fuente ve una impedancia acoplada, y la salida del transductor ve una carga acoplada, figura B.3. Entonces la fuente entrega su potencia de señal *disponible* S_{ai}, y la potencia de la señal de salida es

$$S_{ao} = \mathcal{G}_a S_{ai}$$

donde $\mathcal{G}_a$ es la *ganancia de potencia disponible* del dispositivo. De manera similar, la potencia de ruido de salida es

$$N_{ao} = \mathcal{G}_a N_{ai} + N_{ax}$$

siendo N_{ax} la potencia disponible de ruido excedente en la salida. Si el ruido de entrada es blanco y se representa por una temperatura de ruido $\mathcal{T}_i$, entonces

$$N_{ai} = k\mathcal{T}_i B_N$$

es el ruido de la fuente disponible en el ancho de banda equivalente B_N.

Combinando estos términos se tiene la relación señal a ruido de salida como

$$\left(\frac{S}{N}\right)_o = \frac{S_{ao}}{N_{ao}} = \frac{\mathcal{G}_a S_{ai}}{\mathcal{G}_a k\mathcal{T}_i B_N + N_{ax}} = \frac{S_{ai}}{k\mathcal{T}_i B_N + N_{ax}/\mathcal{G}_a}$$

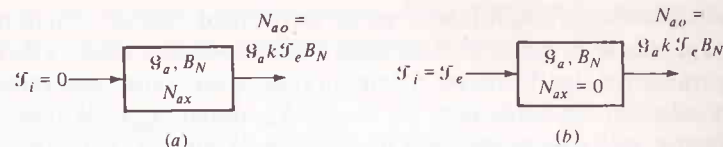



FIGURA B.4
Un amplificador ruidoso con $\mathcal{T}_i$ = 0 y su equivalente sin ruido con $\mathcal{T}_i = \mathcal{T}_e$.

y tomando la relación señal a ruido de entrada como $(S/N)_i = S_{ai}/k\mathcal{T}_i B_N$, se tiene

$$\left(\frac{S}{N}\right)_o = \frac{(S/N)_i}{1 + (N_{ax}/\mathcal{G}_a k\mathcal{T}_i B_N)} \leq \left(\frac{S}{N}\right)_i \tag{3}$$

La ecuación (3) muestra que las relaciones señal a ruido pueden ser casi iguales, no obstante, el ruido excedente, siempre y cuando $N_{ax} \ll \mathcal{G}_a k\mathcal{T}_i B_N$. Es obvio que lo que cuenta no es el valor absoluto de la potencia de ruido excedente sino su valor en relación con el ruido de la fuente, una conclusión que en realidad tiene sentido.

Llevando este enfoque más lejos, se observa que $N_{ax}/\mathcal{G}_a kB_N$ depende sólo de los parámetros del transductor de dos puertas y que tiene dimensiones de temperatura. Por lo tanto, se define a la *temperatura efectiva del ruido de entrada* (conocida también como *temperatura del amplificador*) por

$$\mathcal{T}_e \triangleq \frac{N_{ax}}{\mathcal{G}_a kB_N} \tag{4}$$

En la figura B.4 se ilustra el significado de $\mathcal{T}_e$; en esta figura se muestra que un amplificador ruidoso con $\mathcal{T}_i = 0$ se puede sustituir con un amplificador sin ruido con $\mathcal{T}_i = \mathcal{T}_e$. En pocas palabras, $\mathcal{T}_e$ es una medica del ruido *con relación a la entrada*. Si el dispositivo es sin ruido.

Sustituyendo la ecuación (4) en la (3) se tiene

$$\left(\frac{S}{N}\right)_o = \frac{(S/N)_i}{1 + \mathcal{T}_e/\mathcal{T}_i} \tag{5a}$$

$$= \frac{\mathcal{G}_a S_i}{\mathcal{G}_a k(\mathcal{T}_i + \mathcal{T}_e)B_N} \tag{5b}$$

La ecuación (5*b*) es particularmente informativa, pues dice que bajo condiciones de acoplamiento, el ruido de salida disponible es

$$N_{ao} = \mathcal{G}_a k(\mathcal{T}_i + \mathcal{T}_e)B_N \tag{6}$$

y la temperatura de ruido de salida es $\mathcal{G}_a(\mathcal{T}_i + \mathcal{T}_e)$.

Si las impedancias no están acopladas, todas las potencias deben ser *menores* que las potencias disponibles, quedando reducidas por un factor de desacoplamiento. No obstante, las ecuaciones (3) y (5) son aún válidas, ya que son *relaciones* de potencia tomadas en puntos específicos, y el factor de desacoplamiento se cancela. Así,

aún cuando un amplificador no sea en realidad operado con impedancias acopladas, su ganancia de potencia disponible y su temperatura efectiva de ruido de entrada son parámetros significativos. Como un regalo, éstos son relativamente fáciles de medir o calcular. En particular, $\mathscr{G}_a = S_{ao}/S_{ai}$, donde S_{ai} es la potencia disponible de la fuente, esté o no acoplada la impedancia de entrada, mientras que S_{ao} es la potencia que sería entregada a una carga aceptada de salida cuando la fuente se conecta a la entrada.

Figura de ruido

La temperatura efectiva del ruido de entrada es muy útil en la descripción de los amplificadores de bajo ruido, dispositivos con $\mathscr{T}_e \ll \mathscr{T}_i$. Pero cuando el ruido excedente es grande, la *figura de ruido integrado F* demuestra ser más conveniente.

La figura de ruido se define como la potencia de ruido de salida real dividida por la potencia de ruido de salida si el transductor de dos puertas fuera sin ruido, estando la fuente a la temperatura ambiente $\mathscr{T}_o$. La definición es una relación de potencia y se le puede escribir en términos de potencias disponibles como

$$F \triangleq \frac{N_{ao}}{\mathscr{G}_a k \mathscr{T}_0 B_N} \tag{7}$$

Ahora, con $\mathscr{T}_i = \mathscr{T}_0$, $N_{ao} = \mathscr{G}_a k(\mathscr{T}_0 + \mathscr{T}_e)\ B_N$ y por ello

$$F = \frac{\mathscr{T}_0 + \mathscr{T}_e}{\mathscr{T}_0} = 1 + \frac{\mathscr{T}_e}{\mathscr{T}_0} \tag{8a}$$

$$= 1 + \frac{N_{ax}}{\mathscr{G}_a k \mathscr{T}_0 B_N} \tag{8b}$$

Debe ser aparente que, en general, $F \geq 1$ (el ruido real de salida no puede ser menor que el ruido de la fuente amplificado) y $F = 1$ para un amplificador *sin ruido*.

Para establecer la relación entre las relaciones señal a ruido de entrada y salida, escribiendo la ecuación (8*a*) como

$$\mathscr{T}_e = (F-1)\mathscr{T}_0 \tag{9}$$

y sustituyendo en la ecuación (5*a*) se obtiene

$$\left(\frac{S}{N}\right)_o = \frac{(S/N)_i}{1 + (F-1)(\mathscr{T}_0/\mathscr{T}_i)} \tag{10}$$

Si, *y sólo si*, la temperatura de la fuente es la ambiente, la ecuación (10) se reduce al resultado simple

$$\left(\frac{S}{N}\right)_O = \frac{1}{F}\left(\frac{S}{N}\right)_i$$

Antes del advenimiento de los diseños de bajo ruido, muchas temperaturas de fuente fueron en esencia $\mathscr{T}_o$ (el ruido provino de la resistencia térmica), por lo que la última

expresión se podría emplear en forma indiscriminada. Para los sistemas modernos, en los que el ruido es un factor crucial, se debe emplear la ecuación (10)

Atenuadores y líneas de transmisión

Aunque se han estado considerando los transductores de dos puertas como amplificadores, los resultados anteriores se aplican por igual a los atenuadores y a otros dispositivos que se caracterizan más por presentar *pérdida* de potencia que ganancia. Puesto que tales dispositivos son casi siempre de acoplamiento de impedancia, la pérdida es $\mathscr{L} = 1/\mathscr{G}_a$.

Considérese un transductor de dos puertas compuesto por completo de elementos *resistivos* en equilibrio térmico a la temperatura ambiente $\mathscr{T}_o$. (Este es un buen modelo para líneas de transmisión sin pérdidas y guías de onda, así como redes reales de resistores.) Suponiendo que la temperatura de la fuente es también $\mathscr{T}_o$ y todas las impedancias están acopladas, observando desde las terminales de salida no se ve otra cosa sino una resistencia térmica y un ancho de banda de ruido. Por lo tanto, la potencia de ruido de salida disponible es $N_{ao} = k\mathscr{T}_o B_N$. Pero de la ecuación (6), $N_{ao} = \mathscr{G}_a k(\mathscr{T}_i + \mathscr{T}_e)\ B_N$ para *cualquier* transductor de dos puertas con impedancias acopladas. El dispositivo con pérdidas tiene $\mathscr{G}_a = 1/\mathscr{L}$ y $\mathscr{T}_i = \mathscr{T}_o$, por lo que $\mathscr{T}_e = (\mathscr{L}-1)\ \mathscr{T}_o$, independiente del ancho de banda o de la temperatura de la fuente.

Generalizando, si un transductor de dos puertas con pérdidas está a cualquier temperatura $\mathscr{T}$, su temperatura de ruido efectiva es

$$\mathscr{T}_e = (\mathscr{L}-1)\mathscr{T} \tag{11}$$

y empleando la ecuación (8*a*)

$$F = 1 + (\mathscr{L}-1)\frac{\mathscr{T}}{\mathscr{T}_0} \tag{12}$$

Muy a menudo, $\mathscr{T} = \mathscr{T}_o$, por lo que la ecuación (12) se reduce simplemente a $F = \mathscr{L}$.

Ejemplo B.1

Una técnica sencilla de laboratorio para la medición de $\mathscr{T}$ o de F requiere una resistencia térmica "caliente" y un medidor de potencia relativa, ambos con acoplamiento de impedancia al dispositivo en cuestión, más un termómetro. El procedimiento es como sigue: Primero, conéctese el resistor (a temperatura ambiente) a la entrada del transductor de dos puertas y regístrese la indicación de potencia de salida N_1. Sea C_m la calibración constante del medidor; así

$$N_1 = C_m N_{ao} = C_m \mathscr{G}_a k(\mathscr{T}_0 + \mathscr{T}_e)B_N$$

Segundo, caliéntese el resistor hasta que la lectura de potencia de salida sea *el doble* y regístrese la temperatura del resistor $\mathscr{T}_R$. La nueva lectura es

$$N_2 = C_m \mathscr{G}_a k(\mathscr{T}_R + \mathscr{T}_e)B_N$$

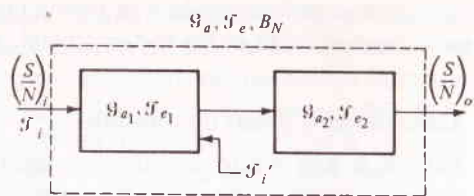



FIGURA B.5
Dos amplificadores ruidosos en cascada.

Ahora $N_2 = 2N_1$, por lo que $C_m\mathcal{G}_a k(\mathcal{T}_R + \mathcal{T}_e)B_N = 2C_m\mathcal{G}_a k(\mathcal{T}_o + \mathcal{T}_e)B_N$, y en consecuencia

$$\mathcal{T}_e = \mathcal{T}_R - 2\mathcal{T}_0 \qquad F = 1 + \frac{\mathcal{T}_e}{\mathcal{T}_0} = \frac{\mathcal{T}_R}{\mathcal{T}_0} - 1 \tag{13}$$

Obsérvese que no se tiene que encontrar a $\mathcal{G}_a$ o a B_N, ni se necesita la constante de calibración C_m. Sin embargo, el resistor estará en el lado caliente, puesto que debe alcanzar un mínimo de ¡2 $\mathcal{T}_o$ = 585°F!

B.2 CÁLCULOS DE RUIDO DEL SISTEMA

Pónganse juntos un determinado número de transductores de dos puertas, representado a la unidad de predetección completa, relacionándose con ello a $(S/N)_R$ con $(S/N)_S$, la relación señal a ruido en las terminales de la antena. Para este propósito se encontrará la temperatura de ruido *total* del receptor o figura de ruido en términos de las propiedades de las etapas individuales.

Considérense dos amplificadores en cascada, como en la figura B.5, suponiendo de nuevo el caso de impedancias acopladas. Es obvio que la ganancia de potencia disponible de la combinación es $\mathcal{G}_a = \mathcal{G}_{a_1}\mathcal{G}_{a_2}$. Sin embargo, la temperatura de ruido total $\mathcal{T}_e$ requiere mayor conocimiento.

Se logra una aproximación examinando la situación en la entrada al segundo amplificador, tratando a todo lo que viene antes como una nueva fuente con $\mathcal{T}_i' = \mathcal{G}_{a_1}(\mathcal{T}_i + \mathcal{T}_{e_1})$. Se introduce después en la ecuación (5), (Sec. B.1), y se gira la manivela. Pero un método más instructivo es observar las diferentes fuentes de ruido y determinar su contribución en la salida. En particular, N_{ao} será la suma de los tres términos:

1 Ruido de fuente amplificado:
2 Ruido excedente de la primera etapa, amplificado por la segunda etapa:
3 Ruido excedente de la segunda etapa:

Se ha supuesto en forma implícita que ambas etapas ven ruido blanco sobre sus pasabandas y que B_N es el ancho de banda total de ruido. En realidad, un buen diseño sugiere que la etapa final debe tener el ancho de banda más pequeño posible para reducir al mínimo el ruido innecesario en el demodulador. Así, ya como normas, se puede tomar a B_N como el ancho de banda de ruido de la última etapa.

Volviendo al cálculo, se suman los tres términos y factorizando se tiene

$$N_{ao} = \mathcal{G}_{a_1}\mathcal{G}_{a_2}k\left[\mathcal{T}_i + \mathcal{T}_{e_1} + \frac{\mathcal{T}_{e_2}}{\mathcal{G}_{a_1}}\right]B_N$$

Pero, por definición de $\mathcal{G}_a$ y $\mathcal{T}_e$ para el caso de en cascada, $N_{ao} = \mathcal{G}_a k(\mathcal{T}_i + \mathcal{T}_e)B_N$; por lo que

$$\mathcal{T}_e = \mathcal{T}_{e_1} + \frac{\mathcal{T}_{e_2}}{\mathcal{G}_{a_1}}$$

y en consecuencia

$$F = F_1 + \frac{F_2 - 1}{\mathcal{G}_{a_1}}$$

Repitiendo este procedimiento con tres o más transductores de dos puertas en cascada, se obtiene para la figura de ruido y la temperatura de entrada efectiva total

$$\mathcal{T}_e = \mathcal{T}_{e_1} + \frac{\mathcal{T}_{e_2}}{\mathcal{G}_{a_1}} + \frac{\mathcal{T}_{e_3}}{\mathcal{G}_{a_1}\mathcal{G}_{a_2}} + \cdots \tag{1}$$

$$F = F_1 + \frac{F_2 - 1}{\mathcal{G}_{a_1}} + \frac{F_3 - 1}{\mathcal{G}_{a_1}\mathcal{G}_{a_2}} + \cdots \tag{2}$$

La última expresión se conoce como la *fórmula de Friis*.

Las ecuaciones (1) y (2) indican que el ruido del receptor se puede controlar en la primera etapa. Supóngase, por ejemplo, que $\mathcal{G}_{a_1} \gg 1$; entonces la temperatura de ruido total es en esencia la de la propia primera etapa, $\mathcal{T}_e \approx \mathcal{T}_{e_1}$. La explicación física es que con una ganancia grande en la primera etapa, el ruido de fuente amplificado (más N_{ax_1}) será mucho mayor que cualquier ruido que se agregue en las siguientes etapas. Por otra parte, si la primera etapa es un *atenuador* de pérdida $\mathcal{L} = 1/\mathcal{G}_{a_1} > 1$ —tal como una línea de transmisión con pérdidas— este efecto es doblemente molesto, puesto que $\mathcal{T}_{e_1} = (\mathcal{L} - 1)\mathcal{T}$ y

$$\mathcal{T}_e = (\mathcal{L} - 1)\mathcal{T} + \mathcal{L}\mathcal{T}_{e_2} + \cdots \tag{3}$$

Esta situación se debe evitar hasta donde sea posible.

Es obvio que las primeras etapas son las de más importancia en el diseño de receptores de comunicaciones de bajo ruido. En particular, la primera etapa debe tener una temperatura de ruido pequeña y una ganancia razonablemente alta. Con una buena "sección de entrada" o preamplificador, las siguientes etapas sirven sólo como amplificación adicional y filtraje, amplificando tanto a la señal como al ruido sin un apreciable cambio en la relación.

Para completar el cálculo, supóngase que la unidad de predetección tiene una temperatura de ruido de entrada efectiva total $\mathcal{T}_e$ y un ancho de banda de ruido

$S_S, \mathcal{T}_S$ → $\mathcal{G}_a, \mathcal{T}_e, B_T$ → $\left(\frac{S}{N}\right)_R$ → Detector → $\left(\frac{S}{N}\right)_D$

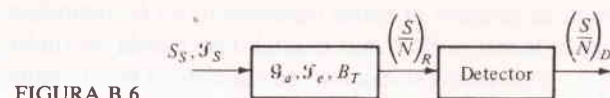



FIGURA B.6

igual en forma aproximada al ancho de banda de la señal modulada, $B_N \approx B_T$. Sea $S_s = S_T/\mathcal{L}$ la potencia de la señal en los terminales de la antena, donde $\mathcal{L}$ es la pérdida de transmisión, y sea $\mathcal{T}_s$ la temperatura de ruido (*temperatura de antena*), figura B.6. (Las antenas receptoras son por lo general de impedancia acoplada en sus terminales, por lo que S_s es la potencia *disponible*.) Así, la relación señal a ruido de la predetección es

$$\left(\frac{S}{N}\right)_R = \frac{\mathcal{G}_a S_S}{\mathcal{G}_a k(\mathcal{T}_S + \mathcal{T}_e)B_T} = \frac{S_T}{\mathcal{L}k\mathcal{T}_N B_T} \tag{4}$$

donde $\mathcal{T}_N = \mathcal{T}_S + \mathcal{T}_e$ es la *temperatura del sistema* referida a las terminales de la antena. Anteriormente se escribio $(S/N)_R = S_R/\eta B_T$. Por comparación con la ecuación (4) se desprende que

$$S_R = \frac{\mathcal{G}_R S_T}{\mathcal{L}} \qquad \eta = \mathcal{G}_R k\mathcal{T}_N$$

donde $\mathcal{G}_R$ es la ganancia de potencia *real* de la unidad de predetección.

Ejemplo B.2

En lugares inapropiados a veces es necesario colocar una antena receptora de televisión sobre un mástil elevado. La antena y el receptor se conectan por medio de un cable largo con las pérdidas del caso. Para contrarrestar los efectos del cable, se puede montar en la antena un preamplificador como se muestra en la figura B.7. Los parámetros del sistema están dados en decibeles, lo cual es normal en la práctica, pero se deben convertir antes de emplearse en las ecuaciones (1) o (2).

Dando valores en la ecuación (2), con $F_2 = \mathcal{L} = 2$, se tiene, para la figura de ruido total

$$F = 4 + \frac{2-1}{100} + \frac{20-1}{100 \times \frac{1}{2}} \doteq 4.39 = 6.4 \text{ dB}$$

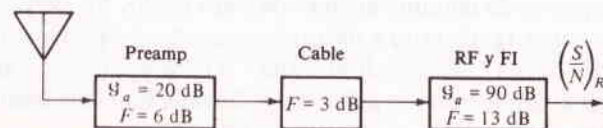



FIGURA B.7
Un receptor con preamplificación en las terminales de antena.

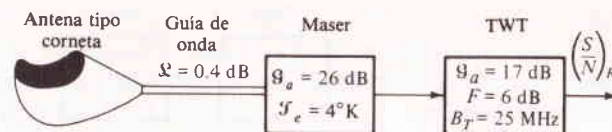



FIGURA B.8
Un receptor de bajo ruido para una estación terrena para satélites.

lo cual demuestra que la generación de ruido es la que se desarrolla en el preamplificador solo. (Nótese que en estos cálculos no intervienen las ganancias en RF y FI.) Así, si la temperatura de la fuente es $\mathcal{T}_s = \mathcal{T}_o$, la $(S/N)_R$ es 6.4 dB menor que la relación señal a ruido en los terminales de la antena.

Si se omite el preamplificador, $F = 2 + 2(20 - 1) = 40$, y el ruido de salida es mayor unas 10 veces respecto de lo anterior.

Ejemplo B.3

Para efectos de las normas de la era espacial, en el ejemplo anterior se tiene un receptor muy ruidoso. En los sistemas de bajo ruido, tales como la estación terrena para satélites de la figura B.8, se tiene ruido excedente suficientemente pequeño porque la figura de ruido es difícil de manejar en forma numérica. Así, por lo general, los cálculos se efectúan en términos de temperaturas.

La guía de onda con pérdidas tiene $\mathcal{L} = 0.4$ dB $= 1.10$, por lo que $\mathcal{T}_{e_1} = (1.10 - 1)\mathcal{T}_o = 29$ K. Para el tubo de ondas progresivas (TWT), $\mathcal{T}_{e_3} = (F_3 - 1)\mathcal{T}_o = 870$ K; por supuesto que la temperatura *física de TWT es mucho menor. Así*

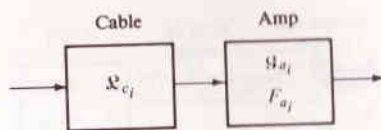
$$\mathcal{T}_e = 29 + 1.1 \times 4 + \frac{1.1 \times 870}{400} \approx 36 \text{ K}$$

Obsérvese que, no obstante las pérdidas pequeñas de la guía de onda, predomina en el sistema su contribución de ruido. El intercambio de posiciones entre el maser y la guía de onda sería deseable en gran manera pero difícil, puesto que la guía de onda es la conexión flexible que permite que la antena de corneta sea orientada. (Con una antena parabólica se puede montar el maser en forma directa al alimentador, eliminándose las pérdidas en la guía de onda y el ruido. Sin embargo, las antenas parabólicas tienden a producir mayores temperaturas de ruido, y de nada sirve el ahorro.)

Bajo condiciones típicas, la temperatura de la antena puede ser $\mathcal{T}_s = 22$ K, dando una temperatura de ruido de sistema de $\mathcal{T}_N = \mathcal{T}_s + \mathcal{T}_e = 58$ K $= 0.2\mathcal{T}_o$. En consecuencia, de la ecuación (4) y empleando $k\mathcal{T}_o \approx 4 \times 10^{-21}$.

$$\left(\frac{S}{N}\right)_R = \frac{S_S}{4 \times 10^{-21}(\mathcal{T}_N/\mathcal{T}_o)B_T} = \frac{S_S}{2 \times 10^{-14}}$$

Una relación señal a ruido de predetección de 20 dB, necesita una potencia de señal disponible en las terminales de antena de $S_s = 10^2 \times 2 \times 10^{-14} = 2\ \mu\mu W$.

FIGURA B.9
La sección i-ésima de un sistema repetidor.

Los sistemas repetidores

Los sistemas repetidores se examinaron en forma breve en la sección 4.1 donde, suponiendo unidades idénticas, se encontró que la relación señal a ruido después de M enlaces era $1/M$ veces la relación señal a ruido para un enlace. Aquí se considera el caso general donde el sistema está formado de M unidades de cable y amplificador, siendo la unidad i-ésima como la mostrada en la figura B.9. La ganancia del amplificador es $\mathscr{G}_{ai}$ y su figura de ruido es F_{ai}; el cable tiene como pérdidas $\mathscr{L}_{ci}$ por lo que + $F_{ei} = \mathscr{L}_{ai}$. Aplicando la ecuación (2), la figura de ruido de la combinación cable/amplificador es

$$F_i = \mathscr{L}_{c_i} + \frac{F_{a_i} - 1}{(1/\mathscr{L}_{c_i})} = \mathscr{L}_{c_i} F_{a_i} \tag{5a}$$

y la unidad tiene una ganancia neta

$$\mathscr{G}_i = \frac{\mathscr{G}_{a_i}}{\mathscr{L}_{c_i}} \tag{5b}$$

Por lo tanto, para M, unidades en cascada,

$$F = \mathscr{L}_{c_1} F_{a_1} + \left(\frac{\mathscr{L}_{c_1}}{\mathscr{G}_{a_1}}\right)(\mathscr{L}_{c_2} F_{a_2} - 1) + \cdots + \frac{\mathscr{L}_{c_1}\mathscr{L}_{c_2}\cdots\mathscr{L}_{c_{M-1}}}{\mathscr{G}_{a_1}\mathscr{G}_{a_2}\cdots\mathscr{G}_{a_{M-1}}}(\mathscr{L}_{c_M} F_{a_M} - 1) \tag{6}$$

y

$$\left(\frac{S}{N}\right)_R = \frac{1}{F}\left(\frac{S}{N}\right)_S \tag{7}$$

puesto que el ruido de entrada o de fuente es casi siempre ruido térmico en $\mathscr{T}_o$.

Para el caso especial pero importante de unidades cercanamente idénticas, donde $\mathscr{L}_{c_1} F_{a_1} = \mathscr{L}_{c_2} F_{a_2} = \ldots$, y una ganancia neta por unidad $\mathscr{G}_i = \mathscr{G}_{a_i}/\mathscr{L}_{c_i} = 1$, la figura de ruido resulta ser

$$\begin{aligned} F &= M\mathscr{L}_{c_1} F_{a_1} - (M-1) \\ &\approx M\mathscr{L}_{c_1} F_{a_1} \qquad \mathscr{L}_{c_1} F_{a_1} \gg 1 \end{aligned} \tag{8a}$$

Por lo que,

$$\left(\frac{S}{N}\right)_R \approx \frac{1}{M\mathscr{L}_{c_1} F_{a_1}}\left(\frac{S}{N}\right)_S = \frac{1}{M}\left(\frac{S}{N}\right)_1 \tag{8b}$$

donde $(S/N)_1 = (1/\mathscr{L}_{e_1} F_{a_1})(S/N)_s$ es la relación señal a ruido al final del primer salto. La ecuación (8a) es la base para la frase *de batalla* empleada por los ingenieros de sistemas telefónicos: "duplicando el número de repetidores se incrementa el factor de ruido en 3 dB".

B.3 PROBLEMAS

B.1 (Secc. B.1) A un amplificador con $\mathscr{G}_a = 50$ dB y $B_N = 20$ kHz se le encontró que tiene $N_{ao} = 10^{10}\, k\mathscr{T}_o$ cuando $\mathscr{T}_i = \mathscr{T}_o$. Encuéntrese $\mathscr{T}_e$ y calcúlese N_{ao} cuando $\mathscr{T}_i = 2\mathscr{T}_o$. (*Respuesta*: $4\mathscr{T}_o$, 48 $\mu\mu W$.)

B.2 (Secc. B.1) Cuando en la entrada de cierto amplificador cambia la temperatura de ruido de $\mathscr{T}_o$ a $2\mathscr{T}_o$, la potencia de ruido de salida disponible se incrementa en un tercio. Encuéntrense $\mathscr{T}_{e_{E1}}$ y F.

B.3 (Secc. B.1) El acoplamiento de impedancias entre una antena de 300 Ω y un amplificador de 50 Ω se puede aproximar poniendo una resistencia de 300 Ω en serie con la antena y una resistencia de 50 Ω a través de las terminales de entrada del amplificador. Calcúlese la figura de ruido de esta red resistiva. (*Sugerencia*: Hágase el análisis considerando el valor de $\mathscr{G}_a$ que se aplica en la ecuación (6).)

B.4 (Secc. B.1) El *diodo generador de ruido* descrito en el problema 3.40 a menudo se emplea como el "resistor caliente" en mediciones de ruido como en el ejemplo B.1. La corriente del diodo no cuenta en el primer paso y después, en el segundo, se ajusta hasta que la potencia de salida se duplica. Demuéstrese que $F = eI_bR/2k\mathscr{T}_o \approx 20 T_b R$. Por lo tanto, si $R = 50$ Ω, entonces F es igual a I_b en miliamperes.

B.5★ (Secc. B.1) Se puede emplear un oscilador senoidal en lugar de resistor caliente del ejemplo B.1. Se conecta el oscilador pero desenergizado en el primer paso, con lo que su resistencia interna entrega el ruido de fuente. Para el segundo paso el oscilador se energiza y su potencia de señal S_i se ajusta para doblar la potencia de salida total. Obténgase una expresión para F y analícense las desventajas de este método.

B.6 (Secc. B.2) Dedúzcase la ecuación (2) para dos etapas en cascada, a partir de la ecuación (5b), (Sec. B.1), considerando a S_i y a $\mathscr{T}_i$ como la salida de la primera etapa.

B.7 (Secc. B.2) Dos amplificadores en cascada tienen las siguientes especificaciones: $\mathscr{T}_{e_1} = 3\mathscr{T}_o$, $\mathscr{G}_{a_1} = 10$ dB, $F_2 = 13.2$ dB, $\mathscr{G}_{a_2} = 50$ dB. Si $B_N = 100$ kHz y $\mathscr{T}_i = 10\mathscr{T}_o$, ¿qué valor de S_i se requiere para dar $(S/N)_o = 30$ dB? (*Respuesta*: 6 $\mu\mu W$.)

B.8 (Secc. B.2) Un sistema consiste de un cable, cuya pérdida es 2 dB/kM, seguido de un amplificador con $F = 7$ dB. Si $\mathscr{T}_i = \mathscr{T}_o$ y se desea tener $(S/N)_o \geq 0.05(S/N)_i$, ¿cuál es la máxima longitud posible de la trayectoria?

B.9 (Secc. B.2) Obténgase una expresión similar a la ecuación (6) cuando el amplificador se coloca antes del cable y no después de éste. Simplifíquese el resultado para el caso de unidades idénticas con $\mathscr{G}_{e_i}/\mathscr{L}_{e_i} = 1$ y compárese con la ecuación (8a).

B.10★ (Secc. B.3) La *medición de ruido Haus-Adler* se define como $\mathscr{M} = (F - 1)/(1 - 1/\mathscr{G}_a)$. Demuéstrese que una cascada de dos amplificadores tiene el factor de ruido total más bajo si el primer amplificador tiene el valor más bajo de (*Sugerencia*: Escríbase una expresión para la diferencia $F_{12} - F_{21}$ para las dos configuraciones posibles.)

APÉNDICE C
LOS SISTEMAS DE TELEVISIÓN Y FACSÍMIL

Los sistemas de televisión y facsímil son sistemas de transmisión de *imágenes*; es decir, que el mensaje es un patrón bidimensional, y por lo tanto, una función de dos variables independientes. Este Apéndice trata de los medios para transmitir una imagen como una señal eléctrica, la cual es una función de una sola variable, el tiempo. Se dará especial atención a la televisión, donde las imágenes tienen movimiento y existe una limitación en transmisión de tiempo real. En los sistemas de facsímil sólo se transmiten imágenes fijas, dando al diseñador una mayor libertad.

C.1 LA SEÑAL DE VIDEO

Para empezar con el caso más simple, considérese un patrón de intensidad monocromático (blanco y negro) de movimiento libre $I(h, v)$, donde h y v son las coordenadas horizontal y vertical. La conversión de $I(h, v)$ a una señal $x(t)$ —y viceversa— requiere de un proceso discontinuo de mapeo, tal como la *trama de exploración* que se muestra en el diagrama de la figura C.1. El dispositivo de exploración, que produce un voltaje o una corriente proporcionales en intensidad, principia en el punto A y se mueve con velocidades constantes, pero no iguales en las direcciones horizontal y vertical, siguiendo la trayectoria AB. Así, si s_h y s_v son las velocidades de exploración horizontal y vertical, la salida del explorador es la *señal de video*,

$$x(t) = I(s_h t, s_v t) \qquad (1)$$

FIGURA C.1
Trama de exploración (el espaciamiento entre líneas está exagerado en forma burda). Las líneas continuas constituyen el primer campo; las líneas interrumpidas el segundo.

puesto que $h = s_h t$, etc. Alcanzar el punto B, el punto de exploración regresa en forma rápida a C (el retrazo horizontal) y prosigue de manera similar hacia el punto D, donde finaliza la exploración de facsímil.

En televisión, sin embargo, se debe propiciar el movimiento de la imagen, por lo que el punto retraza en forma vertical hacia E y produce un patrón entrelazado que finaliza en F. El proceso se repite, empezando de nuevo en A. Los dos conjuntos de líneas se conocen como el primero y segundo *campos*; juntos constituyen una imagen completa o *cuadro*. La velocidad de cuadro es lo suficientemente rápida (25 a 30 por segundo) como para crear la ilusión de movimiento continuo, mientras que la velocidad de campo (del doble de la velocidad de cuadro) hace que el parpadeo sea imperceptible para el ojo humano. En consecuencia, la exploración entrelazada permite la velocidad de repetición de imagen más baja posible sin un parpadeo apreciable.

Se hacen dos modificaciones a la señal de video después de la exploración: se le insertan *pulsos de borrado* durante los intervalos de retraso para suprimir las líneas de éste en el tubo de imagen del receptor; y se agregan *pulsos de sincronización* en la

FIGURA C.2
Forma de onda de video para una línea completa.

TABLA C.1 NORMAS DE TELEVISIÓN EN ESTADOS UNIDOS

Relación de aspecto (anchura a altura)	4/3
Total de líneas por cuadro	525
Frecuencia de línea †	15.75 kHz
Tiempo de línea †	63.5 μs
Tiempo de retrazo horizontal †	10 μs
Frecuencia de campo †	60 Hz
Retrazo vertical †	20 líneas por campo
Ancho de banda de video	4.2 MHz
Ancho de banda de transmisión	6.0 MHz
Frecuencia de la portadora de video	54-72, 76-88, 174-216, 470-890 MHz
Frecuencia de la portadora de audio	4.5 Hz arriba de la portadora de video
Desviación de audio en FM	25 kHz

† Valores nominales.

parte superior de los pulsos de borrado para sincronizar a los circuitos de barrido horizontal y vertical del receptor. La figura C.2 muestra la forma de onda para una línea completa, con niveles de amplitud y duraciones correspondientes a las normas de TV de Estados Unidos. En la tabla C.1 se enumeran otros parámetros normas.

El espectro de video

El análisis del espectro de la señal de video en ausencia de movimiento es relativamente fácil con la ayuda de la figura C.3, donde, en vez de la exploración de retrazo,

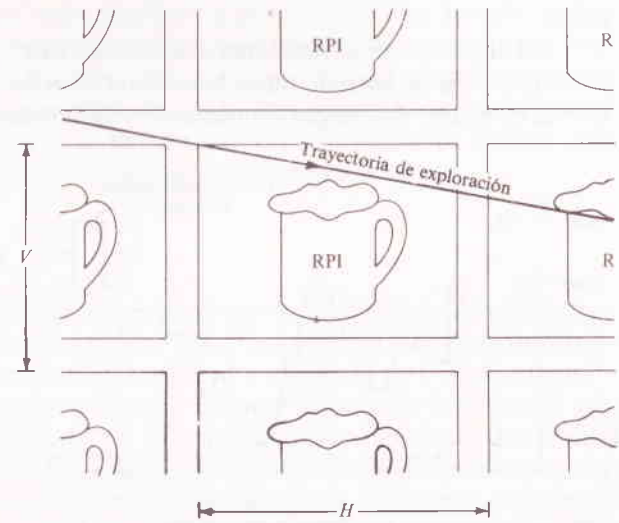



FIGURA C.3
Imagen repetida en forma periódica con trayectoria de exploración no interrumpida.

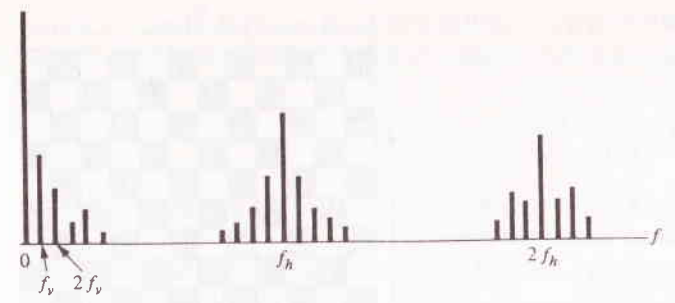



FIGURA C.4
Espectro de video para una imagen fija.

la imagen se repite en forma periódica en ambas direcciones por lo que la trayectoria de exploración equivalente no se interrumpe. Así, cualquier función periódica de dos variables se puede desarrollar como una *serie de Fourier bidimensional* por medio de una extensión directa de la serie unidimensional. Para el caso que se analiza, con H y V como los periodos horizontal y vertical (incluyendo la concesión de retrazo), la intensidad de la imagen es

$$I(h,v) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} c_{mn} \exp\left[j2\pi\left(\frac{mh}{H} + \frac{nv}{V}\right)\right] \tag{2}$$

donde

$$c_{mn} = \frac{1}{HV}\int_0^H \int_0^V I(h,v) \exp\left[-j2\pi\left(\frac{mh}{H} + \frac{nv}{V}\right)\right] dh\, dv \tag{3}$$

Por lo tanto, haciendo

$$f_h = \frac{s_h}{H} \qquad f_v = \frac{s_v}{V}$$

y empleando las ecuaciones (1) y (2)

$$x(t) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} c_{mn} e^{j2\pi(mf_h + nf_v)t} \tag{4}$$

con lo que se tiene una señal periódica del doble conteniendo todas las armónicas de la *frecuencia de línea* f_h y de la *frecuencia de campo* c_{mn}, más sus sumas y diferencias. Puesto que $f_k \gg f_v$, y que c_{mn} por lo general, disminuye conforme aumenta el producto *mn*, el espectro de amplitud tiene la forma que se muestra en la figura C.4, donde las líneas espectrales se agrupan alrededor de las armónicas de f_h y existen grandes intervalos entre los grupos.

La ecuación (4) y la figura C.4 son exactas para una imagen fija, es decir, en sistemas de facsímil. Cuando la imagen tiene movimiento, las líneas espectrales se pierden en grupos continuos alrededor de las armónicas de f_h. Aun así, el espectro permanece "vacío" por lo general, un hecho que se emplea con ventaja en la televisión a color, como se explica en la sección C.3.

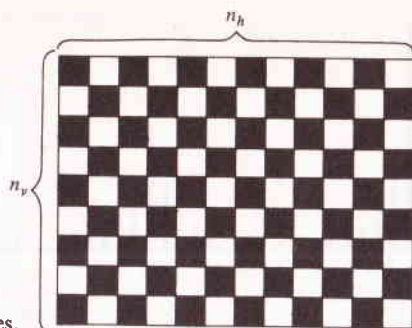



FIGURA C.5
Celdas de resolución horizontales y verticales.

Reducción y ancho de banda

Dos factores básicos que impiden la reproducción perfecta de imágenes: puede haber sólo un número finito de líneas en la trama de exploración, lo cual limita la claridad de la imagen o *resolución* en la dirección vertical; y que la señal de video se debe transmitir con un ancho de banda finito, lo cual limita la resolución horizontal. En forma cuantitativa, se mide la resolución en términos del número máximo de líneas de imagen discretas que se pueden distinguir en cada dirección, digamos n_h y n_v. En otras palabras, la imagen más detallada que se puede definir se considera que es un tablero que tenga n_h columnas y n_v filas (Fig. C.5). Por lo general son de desearse unas resoluciones horizontal y vertical iguales en líneas por distancia unitaria, es decir, N_h/(ancho de imagen) = n_r/(altura de la imagen), o

$$\frac{n_h}{n_v} = \frac{\text{ancho de imagen}}{\text{altura de imagen}} = \mathscr{A} \tag{5}$$

y a $\mathscr{A}$ se le conoce como la *relación de aspecto*.

Es obvio que la resolución vertical está relacionada con el número total de líneas de trama N; además, n_v es igual a N si todas las líneas de exploración son activas en la formación de la imagen (como en facsímil pero no en TV) y la trama se alinea de manera perfecta con las filas de la figura C.5. Los estudios experimentales demuestran que una alineación arbitraria de la trama reduce la resolución efectiva en un factor de alrededor del 70%, conocido como el *factor de Kerr*, por lo que

$$n_v = 0.7(N - N_{vr}) \tag{6}$$

donde N_{vr} es el número de líneas de la trama que se pierden durante el retrazo vertical.

La resolución horizontal se determina por el ancho de banda de banda base B asignado a la señal de video. Si la señal de video es una sinusoide de frecuencia $f_{\text{máx}} = B$, la imagen resultante será una secuencia de puntos luminosos y oscuros en forma alternada, espaciados en medio ciclo en la dirección horizontal. Se desprende entonces que

$$n_h = 2B(T_{\text{línea}} - T_{hr}) \tag{7}$$

donde $T_{\text{línea}}$ es la duración total de una línea y T_{hr} es el tiempo de retrazo horizontal. Resolviendo la ecuación (7) para B y empleando las ecuaciones (5) y (6) se tiene

$$B = \frac{\mathscr{A} n_v}{2(T_{\text{línea}} - T_{hr})} = 0.35\mathscr{A}\,\frac{N - N_{vr}}{T_{\text{línea}} - T_{hr}} \tag{8}$$

Se obtiene una expresión de ancho de banda alterna y más versátil multiplicando ambos miembros de la ecuación (8) por el tiempo de cuadro $T_{\text{cuadro}} = NT_{\text{línea}}$ y se demuestra en forma explícita la resolución vertical. Puesto que $N = n_v/0.7(1 - N_{vr}/N)$, esto da por resultado

$$BT_{\text{cuadro}} = \frac{0.714\mathscr{A} n_v^2}{\left(1 - \frac{N_{vr}}{N}\right)\left(1 - \frac{T_{hr}}{T_{\text{línea}}}\right)} \tag{9}$$

resaltando que el requisito de ancho de banda (o tiempo de cuadro) es proporcional al cuadrado de la resolución.

Ejemplo C.1

En las normas de TV de Estados Unidos,† $N = 525$ y $N_{vr} = 2 \times 20 = 40$ por lo que hay 485 líneas activas. El tiempo de línea es $T_{\text{línea}} = 1/f_h = 6.354$ μs y $T_{hr} = 10$ μs, permitiendo un tiempo de línea activa de 53.5 μs. Por lo tanto, empleando la ecuación (8) con $\mathscr{A} = 4/3$,

$$B = 0.35 \times \frac{4}{3} \times \frac{485}{53.5 \times 10^{-6}} \approx 4.3 \text{ MHz}$$

comparado con el valor específico de 4.2 MHz. Nótese que este ancho de banda es lo bastante grande para asegurar que los pulsos de sincronización de 5μs sufran sólo una leve distorsión.

Ejemplo C.2

A diferencia de la TV, en los sistemas de facsímil no se requiere retraso vertical y el tiempo de retraso horizontal es despreciable. Si se va a transmitir una plana completa de un diario (de 37 por 59 cm) sobre un circuito telefónico ($B \approx 3.2$ kHz) con una resolución de 40 líneas/cm, la ecuación (9) da para el tiempo de transmisión

$$T_{\text{cuadro}} = \frac{0.714(37/59)(40 \times 59)^2}{3.2 \times 10^3} \approx 780 \text{ s} = 13 \text{ min}$$

Cálculos como éste subrayan el hecho de que la transmisión de facsímil de gran resolución requiere un tiempo de transmisión sustancial o de un canal de banda ancha.

† En Europa, $N = 625$ y $f_v = 50$ Hz.

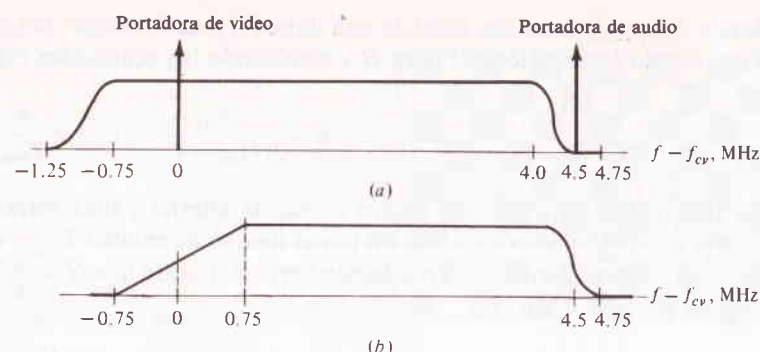



FIGURA C.6
(*a*) Espectro de TV como se transmite; *b*) Forma de la banda lateral residual en el receptor.

C.2 TRANSMISORES Y RECEPTORES DE TV

El ancho de banda grande y el contenido de baja frecuencia significativo de la señal, de video, junto con la sencillez deseada de la detección de envolvente, han conducido a la selección de la VSB + C (como se describió en la Sec. 5.4) para radiodifusión de TV en Estados Unidos. Sin embargo, puesto que la conformación precisa de la banda lateral residual se realiza con más facilidad en el receptor donde los niveles de potencia son pequeños, el espectro real de la señal modulada es como se indica en la figura C.6*a*. La frecuencia de potencia media de la banda lateral superior está alrededor de 4.2 MHz arriba de la portadora de video f_{cv} mientras que la banda lateral inferior tiene un ancho de banda de 1 MHz. La figura C.6*b* muestra la conformación de frecuencia en el receptor.

La señal de audio está modulada en frecuencia sobre una portadora separada $f_{ca} = f_{cv} + f_a, f_a = 45$ MHz, con una desviación de frecuencia $f_\Delta = 25$ kHz. Así, suponiendo un ancho de banda de audio de 10 kHz, $\Delta = 25$, el audio modulado ocupa alrededor de 80 kHz (véase la Sec. 6.3). Los canales de TV están espaciados en 6 MHz, dejando una banda de seguridad de 250 kHz, con frecuencias portadoras asignadas en las bandas de VHF y UHF.

A continuación se proporcionan detalles adicionales acerca de los transmisores y receptores de TV monocromáticos; las modificaciones para TV a color se examinan en la siguiente sección. Los circuitos electrónicos correspondientes se pueden encontrar en la literatura.

Transmisores

En el diagrama a bloques de la figura C.7 se muestran las partes esenciales de un transmisor de TV. El generador de sincronización controla la trama de exploración y

† Por ejemplo, Terman (1955, Cap. 25) o Hansen (1969).

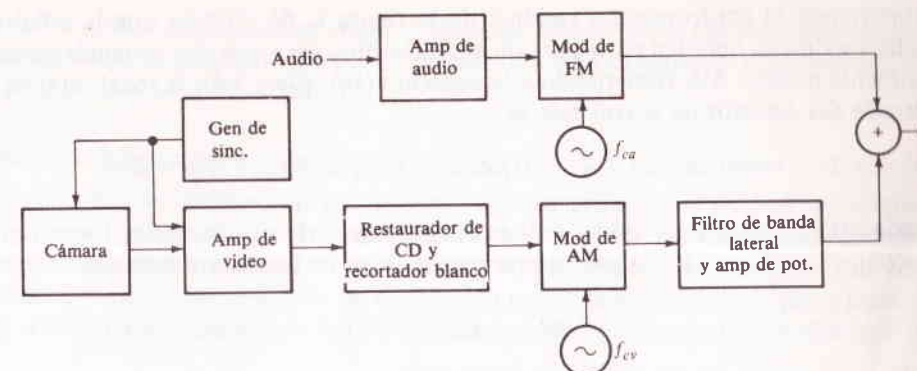



FIGURA C.7
Transmisor de TV.

proporciona los pulsos de borrado y sincronismos para la señal de video. El restaurador de CD y el recortador blanco trabajando juntos aseguran que los niveles de señal de video amplificada estén en las proporciones mostradas en la figura C.2. El modulador de video es del tipo de AM de alto nivel, y el amplificador de potencia remueve la porción inferior de la banda lateral inferior.

La antena tiene una configuración de puente equilibrado de tal manera que las salidas de los transmisores de audio y video sean irradiadas por la misma antena sin que se interfieran entre sí. La potencia transmitida de audio es de 50 a 70% de la potencia de video.

Receptores

Como se indica en la figura C.8, un receptor de TV es del tipo superheterodino (Sec. 5.5). El amplificador principal de FI tiene f_{FI} en el intervalo de 41 a 60 MHz y

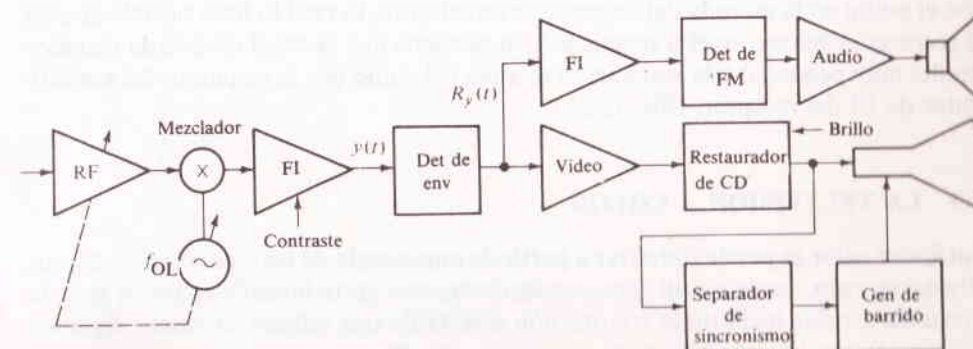



FIGURA C.8
Receptor de TV.

roporciona la conformación residual de la figura C.6*b*. Nótese que la señal de udio modulada también pasa por este amplificador, pero con una ganancia sustancialmente menor. Así, recurriendo a la ecuación (10), (Sec. 5.4), la señal total en la ntrada del detector de envolvente es

$$(t) = A_{cv}[1 + mx(t)] \cos \omega_{cv} t - A_{cv} m\zeta(t) \operatorname{sen} \omega_{cv} t + A_{ca} \cos [(\omega_{cv} + \omega_a)t + \phi(t)] \quad (1)$$

onde $x(t)$ es la señal de video, $\phi(t)$ es el audio de FM, y $\omega_a = 2\pi f\alpha$. Puesto que $|m\zeta(t)| \ll 1$ y $A_{ca} \ll A_{cv}$, la envolvente resultante es en forma aproximada

$$R_y(t) = A_{cv}[1 + mx(t)] + A_{ca} \cos [\omega_a t + \phi(t)] \quad (2)$$

xpresión que da la señal a la salida dèl detector de envolvente.

El amplificador de video tiene un filtro pasabajas para remover la componente e audio de $R_y(t)$ así como un restaurador de CD que en forma electrónica bloquea s pulsos de borrado y con ello restituye el nivel correcto de CD a la señal de video. a señal de video amplificada y con la CD restituida se aplica al tubo de imagen y a n separador de pulsos de sincronismos que proporciona la sincronización para los eneradores de barrido. El control de "brillo" permite el ajuste manual del nivel de D, mientras que el control de "contraste" ajusta la ganancia del amplificador de FI.

La ecuación (2) demuestra que la salida del detector de envolvente también cluye el audio modulado. Esta componente se recoge y amplifica por medio de tro amplificador de FI sintonizado a 4.5 MHz. La detección de la FM y la amplifición dan lugar así a la señal de audio.

Obsérvese que, aunque la señal compuesta de audio y video transmitida es un o de multicanalización por división en frecuencia, no se requiere de conversión de ecuencia por separado para el audio. Esto es así porque la portadora de video aca como un oscilador total para el audio en el proceso de detección de envolvente, arreglo conocido como *sistema de sonido por interportadora* que tiene el aspecto ntajoso de que el audio y el video están siempre sintonizados juntos. La operación itosa depende del hecho de que la componente de video es grande en comparación n el audio en la entrada del detector de envolvente, lo cual lo hace posible gracias recortador blanco en el transmisor (ello previene que la señal modulada de video sulte muy pequeña) y la atenuación relativa del audio por la respuesta del amplifidor de FI del receptor, (Fig. C.6*b*).

3 LA TELEVISIÓN A COLOR

alquier color se puede sintetizar a partir de una mezcla de los tres colores aditivos marios, rojo, verde y azul. Por consiguiente, una aproximación rigurosa para la evisión a color incluiría la transmisión directa de tres señales de video, digamos $t), x_G(t)$, y $x_B(t)$, una para cada color primario. Pero, además de los requisitos de nento del ancho de banda, este método no sería compatible con los sistemas mocromáticos existentes. Una señal de TV de color compatible en forma completa y

que se adapta al canal monocromático, se desarrolló en 1954, aprovechando ciertas características de percepción del color por el hombre. Los aspectos sobresalientes de este sistema se bosquejan aquí.†

Señales de luminancia y crominancia

Por principio de cuentas, las tres señales de color primario se pueden representar únicamente por medio de otras tres señales cualesquiera que sean combinaciones lineales independientes de $x_B(t)$, $x_G(t)$, y $x_B(t)$. Y, con una elección apropiada de coeficientes, se puede hacer una de las combinaciones en la misma forma que la intensidad o señal de *luminancia* de la TV monocromática. En particular, resulta que si

$$x_Y(t) = 0.30x_R(t) + 0.59x_G(t) + 0.11x_B(t) \quad (1a)$$

entonces $x_Y(t)$ es idéntica de manera virtual a la señal de video convencional simbolizada previamente por $x(t)$. Las dos señales sobrantes, conocidas como señales de *crominancia*, se toman como

$$x_I(t) = 0.60x_R(t) - 0.28x_G(t) - 0.32x_B(t) \quad (1b)$$

$$x_Q(t) = 0.21x_R(t) - 0.52x_G(t) + 0.31x_B(t) \quad (1c)$$

Aquí, las señales de color están normalizadas de tal manera que $0 \leq x_R(t) \leq 1$, etc., por lo que la señal de luminancia nunca es negativa, mientras que las señales de crominancia son bipolares.

La comprensión de las señales de crominancia se enriquece con la introducción del *vector de color*

$$x_C(t) = x_I(t) + jx_Q(t) \quad (2)$$

cuya magnitud $|x_c(t)|$ es la intensidad de color o *saturación* y cuyo ángulo arag$[x_c(t)]$ es la *tonalidad*. La figura C.9 muestra las posiciones vectoriales de los colores primarios saturados, en el

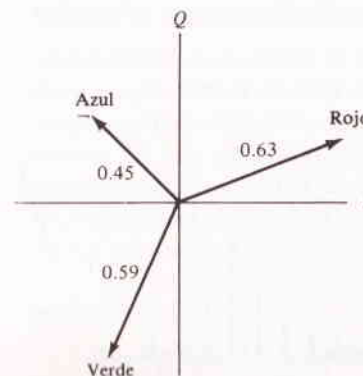



FIGURA C.9
Vectores de colores primarios saturados, en el plano *IQ*.

† Las especificaciones y el análisis completos se dan en Proc. *IRE*, Vol. 42, enero de 1954, el cual está dedicado por completo al tema de la televisión a color.

plano IQ. Un azul-verde (pastel) saturado en forma parcial, por ejemplo, podría tener $x_R = 0$ y $x_B = x_G = 0.5$, por lo que $x_c = -0.300 - j\,0.105, |x_c| = 0.318$, y $\arg[x_c] = -160°$. Puesto que el origen del plano IQ representa la ausencia de color, se puede ver a la señal de luminancia como un vector perpendicular a este plano.

Dado que $x_Y(t)$ sirve como la señal monocromática, se debe asignar el ancho de banda base para proporcionar una resolución horizontal adecuada. En consecuencia, ¡podría parecer que no hay lugar para las señales de crominancia. Recuérdese, sin embargo, que el espectro de $x_Y(t)$ tiene intervalos periódicos entre las armónicas de la frecuencia de línea f_h —y lo mismo es válido para las señales de crominancia. Además, pruebas subjetivas han demostrado que el ojo humano es menos perceptivo a la resolución de crominancia que a la de luminancia, de tal manera que $x_I(t)$ y $x_Q(t)$ se pueden restringir a alrededor de 1.3 MHz y 0.6 MHz, respectivamente, sin una degradación visible significativa de la imagen de color. La combinación de estos factores permite multicanalizar las señales de crominancia en una forma interpolada en el espectro de banda base de la señal de luminancia.

Interpolación de frecuencia y compatibilidad

Las señales de crominancia están multicanalizadas sobre una *subportadora de color* cuya frecuencia cae en forma exacta a la mitad entre la 227ava y la 228ava armónica de f_h, es decir,

$$f_{cc} = \frac{455}{2} f_h \approx 3.6 \text{ MHz} \qquad (3)$$

Por lo tanto, por extensión de la figura C.4, las componentes de frecuencia de luminancia y crominancia se interpolan como se indica en la figura C.10, y hay 0.6 MHz entre F_{cc} y el extremo superior del canal de banda base. La modulación de subportadora se describirá en breve, después de examinar la interpolación de frecuencia y la compatibilidad.

FIGURA C.10
Líneas espectrales de crominancia (interrumpidas) intercaladas con las líneas de luminancia.

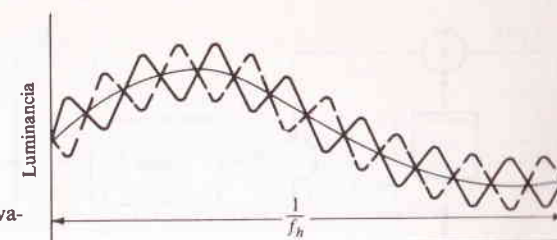



FIGURA C.11
Inversión de fase línea a línea de las variaciones de crominancia a luminancia.

¿Qué sucede cuando se aplica una señal de color a un tubo de imagen monocromático? Nada, en forma sorprendente, hasta donde el espectador puede apreciar. Cierto, la subportadora de color y sus bandas laterales producen variaciones senoidales en la parte superior de la señal de luminancia. Pero como todas estas sinusoides son en forma exacta un múltiplo impar de un medio de la frecuencia de línea, se invierten en fase de línea a línea y de campo a campo (Fig. C.11). Esto produce parpadeo en áreas pequeñas que se promedian sobre el tiempo y el espacio al valor correcto de luminancia y pasa en esencia inadvertido por el espectador.

Por medio de este efecto promediante, la interpolación de frecuencia suministra la señal de color compatible con un receptor monocromático sin modificaciones. Por medio de este efecto también se simplifica el diseño de los receptores de color, puesto que, invirtiendo el argumento anterior, la señal de luminancia no interfiere en forma visible con las señales de crominancia. Existe un problema menor de interferencia causado por la frecuencia diferencia $f_a - f_{cc} = 4\,500 - 3\,579.545 = 920.455$ kHz $= (107/2)f_h$ la cual es una frecuencia "invisible". (Como resultado de este cambio, la frecuencia de campo es en realidad de ¡59.94 Hz en vez de 60 Hz!)

Sistema de multicanalización de crominancia

Cualesquiera dos señales se pueden modular en forma lineal sobre la misma portadora empleando el sistema cuyo diagrama se muestra en la figura C.12. Conocido como *multicanalización de portadora en cuadratura*, este arreglo utiliza el corrimiento de fase en portadora y detección sincrónica para permitir que dos señales ocupen la misma banda de frecuencia; se deja el análisis al lector. La figura C.13 muestra cómo el esquema de portadora en cuadratura se incorpora a un transmisor de TV a color para las señales de crominancia.

Las tres señales de color, primero se matrizan† por la ecuación (1) para formar a $x_Y(t)$, $x_I(t)$, y $x_Q(t)$.

Las señales de crominancia se filtran a pasabajas (con diferentes anchos de banda) y se aplican a los moduladores de subportadora. El filtraje pasabanda subsecuente ge-

† No se muestra la corrección *gamma* no lineal, introducida para compensar la distorsión por brillo de un receptor típico.

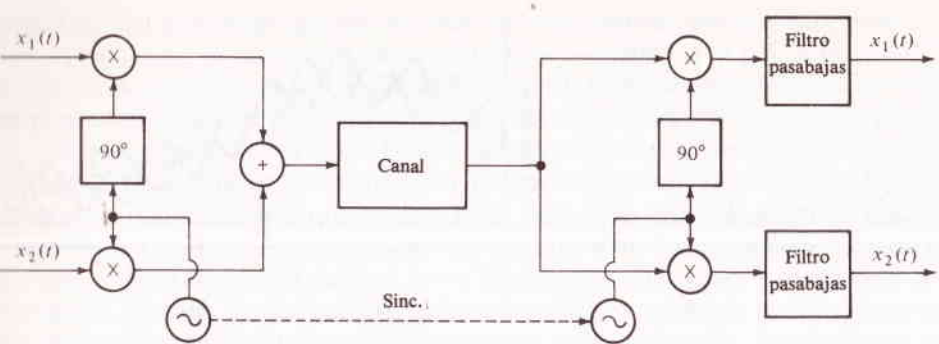



FIGURA C.12
Multicanalización de portadora en cuadratura.

nera la modulación de doble banda lateral convencional para el canal Q y de banda lateral residual modificada para el canal I —es decir, doble banda lateral para frecuencias de banda base de $x_I(t)$ por abajo de 0.6 MHz y banda lateral única inferior para $0.6 < |f| < 1.3$ MHz. Lo último mantiene a las señales de crominancia moduladas tan altas como es posible en el espectro de banda base, confinándose con ello el parpadeo a áreas pequeñas, mientras que sólo se permite el suficiente ancho de banda para una apropiada resolución de $x_I(t)$. No se puede emplear la supresión total de banda lateral obedeciendo al contenido de baja frecuencia significativo en $x_I(t)$ y $x_Q(t)$.

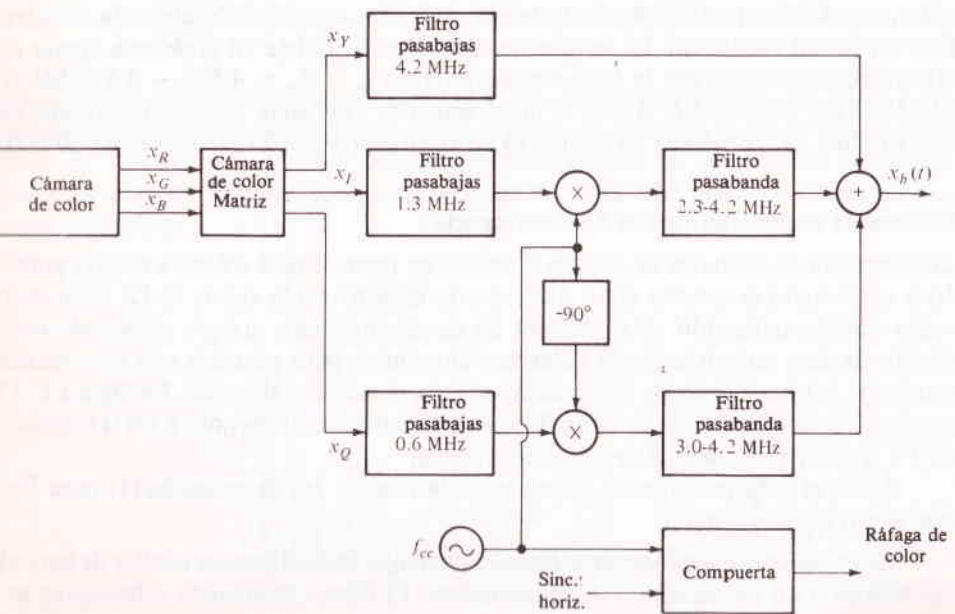



FIGURA C.13
Sistema de modulación de subportadora de color.

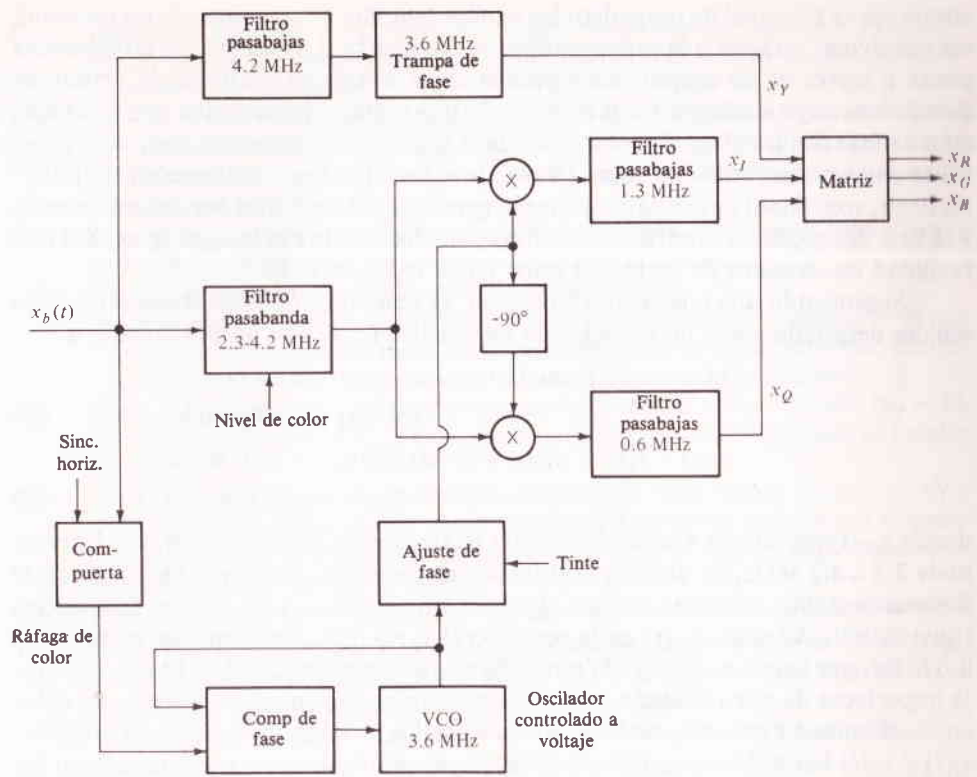



FIGURA C.14
Sistema de demodulación de color.

Incluyendo a $x_Y(t)$, la señal completa de banda base resulta ser

$$x_b(t) = x_Y(t) + x_Q(t)\operatorname{sen}\omega_{cc} t + x_I(t)\cos\omega_{cc} t + \hat{x}_{IH}(t)\operatorname{sen}\omega_{cc} t \qquad (4)$$

donde $\hat{x}_{IH}(t)$ es la transformada de Hilbert de la porción de alta frecuencia de $x_I(t)$ y que cuenta para las bandas laterales asimétricas. Esta señal de banda base toma el lugar de la señal de video monocromática en la figura C.7. En forma adicional, una parte de 8 ciclos de la subportadora de color conocida como *ráfaga de color* se pone sobre la porción posterior o "pórtico posterior" de los pulsos de borrado para efectos de sincronización.

La demulticanalización se efectúa en un receptor de TV de color después del detector de envolvente, como se muestra en la figura C.14. Puesto que la señal de luminancia está aquí en banda base, no se requiere mayor procesamiento salvo para amplificación y de una trampa de 3.6 MHz o filtro de supresión para eliminar la

omponente principal de parpadeo; las bandas laterales de crominancia no necesitan er removidas, gracias a la interpolación de frecuencia. Las señales de crominancia asan a través de un amplificador pasabanda y se aplican a un par de detectores incrónicos cuyo oscilador local es controlado a voltaje sincronizaco por compara-ión de fase con la ráfaga de color recibida. Los controles manuales que por lo gene-al se encuentran rotulados como "nivel de color" (es decir, saturación) y "tinte" es decir, tonalidad) sirven para ajustar la ganancia del amplificador de crominancia la fase del oscilador controlado a voltaje; su efecto sobre la imagen se explica con acilidad en términos de vector de color y con la figura C.9.

Suponiendo una buena sincronización, se desprende de la ecuación (4) que las eñales detectadas pero no filtradas de los canales I y Q son proporcionales a

$$v_I(t) = x_I(t) + 2x_{YH}(t)\cos\omega_{cc}t + x_I(t)\cos 2\omega_{cc}t + [x_Q(t) + \hat{x}_{IH}(t)]\operatorname{sen}2\omega_{cc}t \qquad (5a)$$

$$v_Q(t) = x_Q(t) + \hat{x}_{IH}(t) + 2x_{YH}(t)\operatorname{sen}\omega_{cc}t + x_I(t)\operatorname{sen}2\omega_{cc}t - [x_Q(t) + \hat{x}_{IH}(t)]\cos 2\omega_{cc}t \qquad (5b)$$

onde $x_{YH}(t)$ representa a las componentes de frecuencia de luminancia en el interva-) de 2.3 a 4.2 MHz. Es obvio que el filtraje de pasabajas removerá los términos de recuencia doble, mientras que los términos que incluyen a $x_{YH}(t)$ son frecuencias invisibles". Además, $x_{IH}(t)$ en la ecuación (5b) no tiene componentes menores de .6 MHz, por lo que es rechazada por el filtro pasabajas del canal Q. (Aquí, el filtra-: imperfecto da por resultado un efecto molesto conocido como diafonía de color n cuadratura.) Por tanto, no haciendo caso de los términos de frecuencia invisible, (t) y $x_Q(t)$ han sido recuperados y se les puede matrizar con $x_Y(t)$ para generar las eñales de color para el tubo de imagen. De manera específica, de la ecuación (1),

$$x_R(t) = x_Y(t) - 0.96x_I(t) + 0.62x_Q(t) \qquad (6a)$$

$$x_G(t) = x_Y(t) - 0.28x_I(t) - 0.64x_Q(t) \qquad (6b)$$

$$x_B(t) = x_Y(t) - 1.10x_I(t) + 1.70x_Q(t) \qquad (6c)$$

i ocurre que la señal recibida es monocromática, entonces las tres señales de color erán iguales y la imagen reproducida será en blanco y negro. A esto se le conoce omo *compatibilidad inversa*.

.4. PROBLEMAS

.1 (Secc. C.1) Explíquense los siguientes dos enunciados:
(a) El número total de líneas por cuadro en un sistema de TV debe ser impar.
(b) La forma de onda que impulsa la trayectoria de exploración en un sistema de TV debe ser de diente de sierra y no triangular o sinoidal.

.2 (Secc. C.1) No haciendo caso del retrazo y de la pequeña pendiente de la trama, descríbase el espectro de video cuando la imagen está formada de barras verticales alternadas blancas y negras de ancho $H/4$. Repítase para barras horizontales de altura $V/4$ y compárense. [*Sugerencia*: Considérese la propia señal de video en vez de emplear la ecuación (4).]

C.3 (Secc. C.1) Encuéntrese $|c_{mn}|$ para una imagen completamente negra ($I = 0$) excepto por un rectángulo blanco centrado ($I = 1.0$) de αH de ancho por βV de alto. (*Respuesta*: $\alpha\beta|\operatorname{senc} m\alpha \operatorname{senc} n\beta|$.)

C.4 (Secc. C.1) Con relación al problema C.3, constrúyase gráfica del espectro de amplitud para efectos de medición si $f \geq 0$ cuando $\alpha = 0.6$ $\beta = 0.4$ y $f_h = 20f_v$. No se tomen en cuenta aquellas líneas para las cuales $|c_{mn}| < 0.1\,|c_{00}|$. Repítase con $\alpha = 0.4$ y $\beta = 0.6$, y explíquense las diferencias.

C.5 (Secc. C.1) La imagen del teléfono televisivo es cuadrada, el tiempo de línea activa es de alrededor de 100 μs, y hay alrededor de 230 líneas activas.
(a) Encuéntrese el requisito de ancho de banda de video.
(b) Si la señal de video se transmite por medio de modulación por codificación de pulsos binario (en realidad modulación por codificación de pulsos coherente diferencial) con $Q = 8$, ¿cuál es el ancho de banda de transmisión mínimo?

C.6 (Secc. C.1) Calcúlense n_v y B para un sistema de TV con $\mathscr{A} = 4/3$, $N = 625$, $N_{vr} = 48$, $f_v = 50$ Hz, y $T_{hr} = 0.16\ T_{\text{linea}}$. Compárese con las normas de TV que rigen en Estados Unidos.

C.7 (Secc. C.1) Un sistema de transmisión de imágenes tiene valores fijos de N_{vr}/N y T_{hr}/T_{linea}. Se especifica además que la resolución debe ser de n_0 *líneas por distancia unitaria* en ambas direcciones. Demuéstrese que $BT_{\text{cuadro}} = 0.714\,HVn_0^2$, lo cual es independiente de la relación de aspecto.

C.8 (Secc. C.2) Empléese un diagrama fasorial para deducir la ecuación (2) a partir de la (1).

C.9 (Secc. C.2) Supóngase que el proceso de exploración en una cámara de TV es tal que su señal de salida es $\bar{x}(t)\bar{N}^t_{t-\Delta}s(\lambda)d\lambda$ donde $x(t)$ es la señal de video deseada y Δ T_{linea}. A esto se le conoce como el *efecto de apertura* horizontal.
(a) Descríbase en forma cualitativa lo que le ocurre a la imagen reproducida.
(b) En forma cuantitativa, diséñese un filtro que compense el efecto.

C.10 (Secc. C.3) Analícese el *sistema multicanal de portadora en cuadratura*, figura C.12. En particular, obténganse expresiones para las señales de salida cuando el oscilador local del receptor tiene un error de fase arbitrario.

C.11 (Secc. C.3) Desarróllense los detalles que conducen a las ecuaciones (5a) y (5b) a partir de la ecuación (4).

C.12 (Secc. C.3) Empleando la figura C.9, descríbase lo que sucede a la imagen de color reproducida cuando: la ganancia del amplificador de crominancia es muy alta o muy baja; la fase del oscilador controlado a voltaje tiene un error de $\pm 90°$ o $180°$.

C.13★ (Secc. C.3) Encuéntrense expresiones equivalentes a las ecuaciones (4) y (5) cuando todos los filtros del canal x_Q (en transmisor y receptor) son los mismos que en el canal x_I. Examínense los resultados.

TABLA A
PARES TRANSFORMADOS DE FOURIER

Definiciones

Transformada $V(f) = \mathscr{F}[v(t)] = \int_{-\infty}^{\infty} v(t)e^{-j2\pi ft}\,dt$

Transformada inversa $v(t) = \mathscr{F}^{-1}[V(f)] = \int_{-\infty}^{\infty} V(f)e^{j2\pi ft}\,df$

Teorema integral

$$\int_{-\infty}^{\infty} v(t)w^*(t)\,dt = \int_{-\infty}^{\infty} V(f)W^*(f)\,df$$

Teoremas

Operación	Función	Transformada
Linealidad	$\alpha v(t) + \beta w(t)$	$\alpha V(f) + \beta W(f)$
Retardo en tiempo	$v(t - t_d)$	$V(f)e^{-j\omega t_d}$
Cambio de escala	$v(at)$	$\frac{1}{\|a\|} V\left(\frac{f}{a}\right)$
Conjugación	$v^*(t)$	$V^*(-f)$
Dualidad	$V(t)$	$v(-f)$
Traslación de frecuencia	$v(t)e^{j\omega_c t}$	$V(f - f_c)$
Modulación	$v(t)\cos(\omega_c t + \theta)$	$\frac{1}{2}[e^{j\theta}V(f - f_c) + e^{-j\theta}V(f + f_c)]$
Diferenciación	$\frac{d^n v(t)}{dt^n}$	$(j2\pi f)^n V(f)$
Integración	$\int_{-\infty}^{t} v(\lambda)\,d\lambda$	$(j2\pi f)^{-1}V(f)$
Convolución	$v * w(t)$	$V(f)W(f)$
Multiplicación	$v(t)w(t)$	$V * W(f)$
Multiplicación por t^n	$t^n v(t)$	$(-j2\pi)^{-1}\frac{d^n V(f)}{df^n}$

Transformadas

Función	$x(t)$	$X(f)$
Rectangular	$\Pi\left(\frac{t}{\tau}\right)$	$\tau \operatorname{senc} f\tau$
Triangular	$\Lambda\left(\frac{t}{\tau}\right)$	$\tau \operatorname{senc}^2 f\tau$
Gaussiana	$e^{-\pi(t/\tau)^2}$	$\tau e^{-\pi(f\tau)^2}$
Exponencial, de un lado	$e^{-t/\tau}u(t)$	$\frac{\tau}{1 + j2\pi f\tau}$
Exponencial, de dos lados	$e^{-\|t\|/\tau}$	$\frac{2\tau}{1 + (2\pi f\tau)^2}$
Senc	$\operatorname{senc} 2Wt$	$\frac{1}{2W}\Pi\left(\frac{f}{2W}\right)$
Constante	1	$\delta(f)$
Fasorial	$e^{j(\omega_c t + \phi)}$	$e^{j\phi}\delta(f - f_c)$
Sinusoidal	$\cos(\omega_c t + \phi)$	$\frac{1}{2}[e^{j\phi}\delta(f - f_c) + e^{-j\phi}\delta(f + f_c)]$
Impulso	$\delta(t - t_d)$	$e^{-j\omega t_d}$
De muestreo	$\sum_{k=-\infty}^{\infty} \delta(t - kT_s)$	$f_s \sum_{n=-\infty}^{\infty} \delta(f - nf_s)$
Signo	$\operatorname{sgn} t$	$-j/\pi f$
De escalón	$u(t)$	$\frac{1}{2}\delta(f) + \frac{1}{j2\pi f}$

TABLA B
RELACIONES MATEMÁTICAS ÚTILES

Algunas de las relaciones matemáticas que se encuentran en este texto, se enumeran a continuación como una referencia conveniente. Sin embargo, esta tabla no pretende ser un sustituto de manuales más completos.

Identidades trigonométricas

$e^{\pm j\theta} = \cos\theta \pm j \text{ ser } \theta$

$\text{cos}\ \theta = \frac{1}{2}(e^{j\theta} + e^{-j\theta}) = \text{sen}(\theta + 90°)$

$\text{sen}\ \theta = \frac{1}{2j}(e^{j\theta} - e^{-j\theta}) = \cos(\theta - 90°)$

$\text{sen}^2\,\theta + \cos^2\theta = 1$

$\cos^2\theta - \text{sen}^2\,\theta = \cos 2\theta$

$\cos^2\theta = \frac{1}{2}(1 + \cos 2\theta)$

$\cos^3\theta = \frac{1}{4}(3\cos\theta + \cos 3\theta)$

$\text{en}^2\,\theta = \frac{1}{2}(1 - \cos 2\theta)$

$\text{en}^3\,\theta = \frac{1}{4}(3\,\text{sen}\,\theta - \text{sen}\,3\theta)$

$\text{sen}(\alpha \pm \beta) = \text{sen}\,\alpha\cos\beta \pm \cos\alpha\,\text{sen}\,\beta$

$\text{cos}(\alpha \pm \beta) = \cos\alpha\cos\beta \mp \text{sen}\,\alpha\,\text{sen}\,\beta$

$\text{an}(\alpha \pm \beta) = \frac{\tan\alpha \pm \tan\beta}{1 \mp \tan\alpha\tan\beta}$

$\text{sen}\,\alpha\,\text{sen}\,\beta = \frac{1}{2}\cos(\alpha - \beta) - \frac{1}{2}\cos(\alpha + \beta)$

$\text{os}\,\alpha\cos\beta = \frac{1}{2}\cos(\alpha - \beta) + \frac{1}{2}\cos(\alpha + \beta)$

$\text{en}\,\alpha\cos\beta = \frac{1}{2}\text{sen}(\alpha - \beta) + \frac{1}{2}\text{sen}(\alpha + \beta)$

$A\cos(\theta + \alpha) + B\cos(\theta + \beta) = C\cos\theta - S\,\text{sen}\,\theta = R\cos(\theta + \phi)$

londe

$C = A\cos\alpha + B\cos\beta$

$S = A\,\text{sen}\,\alpha + B\,\text{sen}\,\beta$

$R = \sqrt{C^2 + S^2} = \sqrt{A^2 + B^2 + 2AB\cos(\alpha - \beta)}$

$\phi = \arctan\frac{S}{C} = \arctan\frac{A\,\text{sen}\,\alpha + B\,\text{sen}\,\beta}{A\cos\alpha + B\cos\beta}$

Desarrollos en serie y aproximaciones

$(1 + x)^n = 1 + nx + \frac{n(n-1)}{2!}x^2 + \cdots \qquad |nx| < 1$

$e^x = 1 + x + \frac{1}{2!}x^2 + \cdots$

$a^x = 1 + x\ln a + \frac{1}{2!}(x\ln a)^2 + \cdots$

$\ln(1 + x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 + \cdots$

$\text{sen}\,x = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \cdots$

$\cos x = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \cdots$

$\tan x = x + \frac{1}{3}x^3 + \frac{2}{15}x^5 + \cdots$

$\text{arcsen}\,x = x + \frac{1}{6}x^3 + \frac{3}{40}x^5 + \cdots$

$$\arctan x = \begin{cases} x - \frac{1}{3}x^3 + \frac{1}{5}x^5 - \cdots & |x| < 1 \\ \frac{\pi}{2} - \frac{1}{x} + \frac{1}{3x^3} - \cdots & x > 1 \end{cases}$$

$\text{senc}\,x = 1 - \frac{1}{3!}(\pi x)^2 + \frac{1}{5!}(\pi x)^4 - \cdots$

Sumas

$\sum_{m=1}^{M} m = \frac{M(M+1)}{2}$

$\sum_{m=1}^{M} m^2 = \frac{M(M+1)(2M+1)}{6}$

$\sum_{m=1}^{M} m^3 = \frac{M^2(M+1)^2}{4}$

$\sum_{m=0}^{M} x^m = \frac{(x^M - 1)}{(x - 1)}$

Integrales definidas

$$\int_0^\infty \frac{x^{m-1}}{1+x^n}\,dx = \frac{\pi/n}{\operatorname{sen}(m\pi/n)} \qquad n>m>0$$

$$\int_0^\infty \frac{\operatorname{sen} x}{x}\,dx = \int_0^\infty \frac{\tan x}{x}\,dx = \frac{\pi}{2}$$

$$\int_0^\infty \frac{\operatorname{sen} x \cos nx}{x}\,dx = \begin{cases} \dfrac{\pi}{2} & n^2<1 \\ \dfrac{\pi}{4} & n^2=1 \\ 0 & n^2>1 \end{cases}$$

$$\int_0^\infty \frac{\operatorname{sen}^2 x}{x^2}\,dx = \frac{\pi}{2}$$

$$\int_0^\infty \frac{\cos nx}{1+x^2}\,dx = \frac{\pi}{2} e^{-|n|}$$

$$\int_0^\infty \operatorname{senc} x\,dx = \int_0^\infty \operatorname{senc}^2 x\,dx = \tfrac{1}{2}$$

$$\int_0^\infty \operatorname{sen} x^2\,dx = \int_0^\infty \cos x^2\,dx = \frac{1}{2}\sqrt{\frac{\pi}{2}}$$

$$\int_0^\infty x^n e^{-ax}\,dx = \frac{n!}{a^{n+1}} \qquad n\ge 1,\ a>0$$

$$\int_0^\infty e^{-a^2x^2}\,dx = \frac{1}{2a}\sqrt{\pi} \qquad a>0$$

$$\int_0^\infty x^2 e^{-x^2}\,dx = \tfrac{1}{4}\sqrt{\pi}$$

$$\int_0^\infty e^{-ax}\cos x\,dx = \frac{a}{1+a^2} \qquad a>0$$

$$\int_0^\infty e^{-ax}\operatorname{sen} x\,dx = \frac{1}{1+a^2} \qquad a>0$$

$$\int_0^\infty e^{-a^2x^2}\cos bx\,dx = \frac{1}{2a}\sqrt{\pi}\, e^{-(b/2a)^2}$$

TABLA C
LA FUNCIÓN SENC

En esta tabla se dan los valores numéricos de senc x = (sen πx)/πx y sus cuadrados para x 0 a 3.9% en incrementos de 0.1.

x	senc x	senc^2 x	x	senc x	senc^2 x
0.0	1.000	1.000	2.0	0.000	0.000
0.1	0.984	0.968	2.1	0.047	0.002
0.2	0.935	0.875	2.2	0.085	0.007
0.3	0.858	0.737	2.3	0.112	0.013
0.4	0.757	0.573	2.4	0.126	0.016
0.5	0.637	0.405	2.5	0.127	0.016
0.6	0.505	0.255	2.6	0.116	0.014
0.7	0.368	0.135	2.7	0.095	0.009
0.8	0.234	0.055	2.8	0.067	0.004
0.9	0.109	0.012	2.9	0.034	0.001
1.0	0.000	0.000	3.0	0.000	0.000
1.1	−0.089	0.008	3.1	−0.032	0.001
1.2	−0.156	0.024	3.2	−0.058	0.003
1.3	−0.198	0.039	3.3	−0.078	0.006
1.4	−0.216	0.047	3.4	−0.089	0.008
1.5	−0.212	0.045	3.5	−0.091	0.008
1.6	−0.189	0.036	3.6	−0.084	0.007
1.7	−0.151	0.023	3.7	−0.070	0.005
1.8	−0.104	0.011	3.8	−0.049	0.002
1.9	−0.052	0.003	3.9	−0.025	0.001

TABLA D
PROBABILIDADES GAUSSIANAS

La probabilidad de que una variada gaussiana con valor medio m y variancia σ^2 tendrá un valor observado mayor que $m + \kappa\sigma$, lo da la función

$$Q(\kappa) \triangleq \frac{1}{\sqrt{2\pi}} \int_{\kappa}^{\infty} e^{-\lambda^2/2}\, d\lambda$$

conocida como el área bajo la cola gaussiana. Así

$$P(X > m + \kappa\sigma) = P(X \leq m - \kappa\sigma) = Q(\kappa)$$
$$P(|X - m| > \kappa\sigma) = 2Q(\kappa)$$
$$P(m < X \leq m + \kappa\sigma) = P(m - \kappa\sigma < X \leq m) = \tfrac{1}{2} - Q(\kappa)$$
$$P(|X - m| \leq \kappa\sigma) = 1 - 2Q(\kappa)$$

Otras funciones relacionadas con $Q(\kappa)$ son como sigue:

$$\text{erf}\, \kappa \triangleq \frac{2}{\sqrt{\pi}} \int_0^{\kappa} e^{-\lambda^2}\, d\lambda = 1 - 2Q(\sqrt{2}\,\kappa)$$

$$\text{erfc}\, \kappa \triangleq \frac{2}{\sqrt{\pi}} \int_{\kappa}^{\infty} e^{-\lambda^2}\, d\lambda = 1 - \text{erf}\, \kappa = 2Q(\sqrt{2}\,\kappa)$$

$$\Phi(\kappa) \triangleq \frac{1}{\sqrt{2\pi}} \int_0^{\kappa} e^{-\lambda^2/2}\, d\lambda = \tfrac{1}{2} - Q(\kappa)$$

Se supone en todo lo anterior que $\kappa \geq 0$. Cuando el argumento de Q es negativo, se puede emplear la relación

$$Q(-\kappa) = 1 - Q(\kappa)$$

De los valores numéricos de $Q\kappa$ se construye la gráfica a continuación para $0 \leq \kappa \leq 7.0$. Para valores mayores de κ, se puede aproximar $Q(\kappa)$ por medio de

$$Q(\kappa) \approx \frac{1}{\sqrt{2\pi}\,\kappa} e^{-\kappa^2/2} \qquad \kappa \gg 1$$

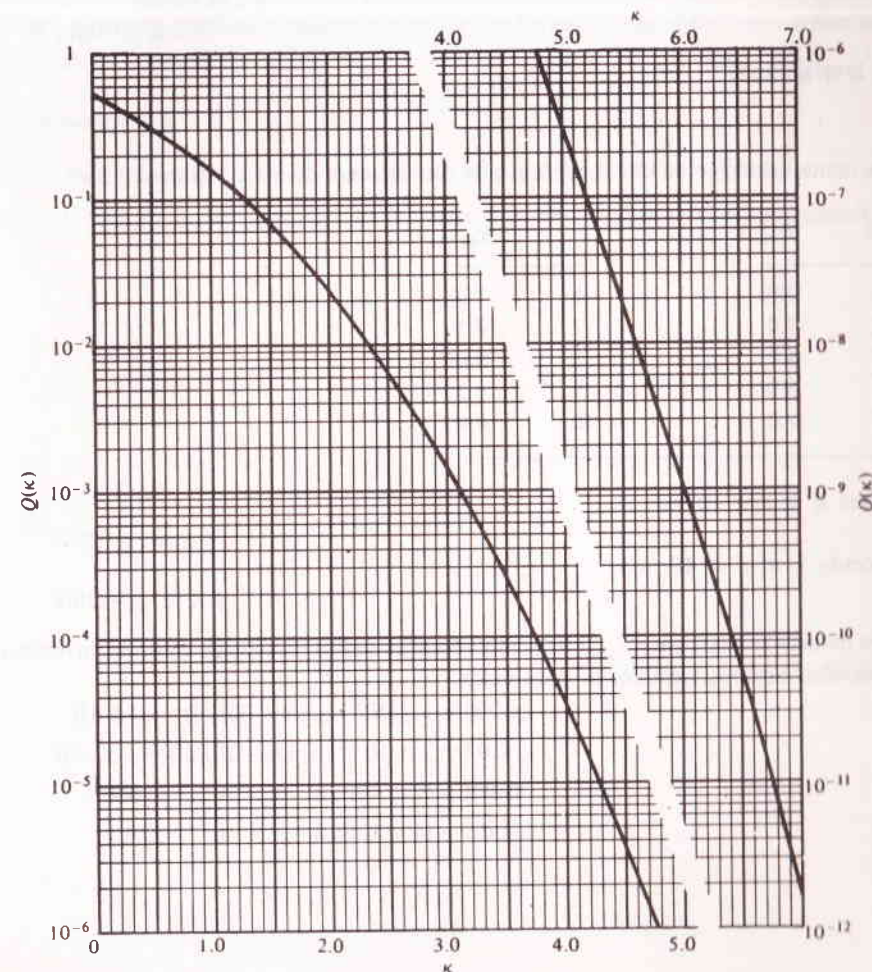



TABLA E
DECIBELES

ingeniería de comunicaciones, los decibeles se emplean casi siempre para *relaciones de po-*
cia. De manera específica, si R es una relación de potencias, entonces

$$R_{dB} \triangleq 10 \log_{10} R$$

nversamente

$$R = 10^{(R_{dB}/10)}$$

continuación se da una tabla breve de conversiones de dB aproximadas para $1 \leq R \leq 10$.

R_{dB}	R	R_{dB}
0.0	6	7.8
3.0	7	8.5
4.8	8	9.0
6.0	9	9.5
7.0	10	10.0

a $R < r < 10$, así

$$R = r \times 10^{\pm N}$$

de $1 < r < 10$; así

$$R_{dB} = r_{dB} \pm 10N$$

manera similar, para $R_{dB} > 10$, empléese la ecuación precedente en orden inverso. A conti-
ción, algunos ejemplos de conversión.

$$6{,}000 = 6 \times 10^3 = (7.8 + 30)\ \text{dB} = 37.8\ \text{dB}$$

$$0.02 = 2 \times 10^{-2} = (3.0 - 20)\ \text{dB} = -17\ \text{dB}$$

$$49\ \text{dB} = (9 + 40)\ \text{dB} = 8 \times 10^4 = 80{,}000$$

$$-4\ \text{dB} = (6 - 10)\ \text{dB} = 4 \times 10^{-1} = 0.4$$

Los decibeles se emplean también para indicar valores absolutos de potencia agregando una tercera letra, es decir, W para watts y mW para miliwatts Por ejemplo:

$$37\ \text{dBW} = 5 \times 10^3\ \text{W} = 5\ \text{kW}$$

$$0\ \text{dBm} = 1\ \text{mW} = 10^{-3}\ \text{W}$$

$$-21\ \text{dBm} = 8 \times 10^{-3}\ \text{mW} = 8\ \mu\text{W}$$

Puesto que el decibel es una unidad logarítmica, los cálculos que incluyen multiplicación, división, potencias y raíces, se pueden desarrollar en forma directa en decibeles, como se ilustra en el siguiente ejemplo:

$$V = \frac{WX^3\sqrt{Y}}{Z}$$

$$V_{dB} = W_{dB} + 3X_{dB} + \tfrac{1}{2}Y_{dB} - Z_{dB}$$

Sin embargo, la suma y la resta, precisan de la conversión de decibeles a valores numéricos.

TABLA F
GLOSARIO DE NOTACIÓN SIMBÓLICA

peraciones		Sección de referencia
*	Complejo conjugado	2.1
e $[H]$, Im $[H]$	Partes reales e imaginarias	2.1
$H\|$	Magnitud o valor absoluto	2.1
g $[H] = \arctan \dfrac{\text{Im}\,[H]}{\text{Re}\,[H]}$	Ángulo de una cantidad compleja	2.1
$(t)\rangle = \lim_{T\to\infty} \dfrac{1}{T}\int_{-T/2}^{T/2} v(t)\,dt$	Promedio de tiempo	2.2
$[v(t)] = \int_{-\infty}^{\infty} v(t)e^{-j2\pi ft}\,dt$	Transformada de Fourier	2.3
$^{-1}[V(f)] = \int_{-\infty}^{\infty} V(f)e^{j2\pi ft}\,df$	Transformada inversa de Fourier	2.3
$w(t) = \int_{-\infty}^{\infty} v(\lambda)w(t-\lambda)\,d\lambda$	Convolución	2.4
$t),w(t)\rangle$	Producto escalar	2.6, A.1
$\| = \langle v(t),v(t)\rangle^{1/2}$	Norma	2.6, A.1
$(\tau) = \langle v(t), w(t-\tau)\rangle$	Correlación cruzada	2.6, 3.5
$\tau) = R_{vv}(\tau)$	Autocorrelación	2.6, 3.5
$f) = \mathscr{F}[R_v(\tau)]$	Densidad espectral	2.6, 3.5
	Valor medio	3.3
$(t)]$	Promedio de ensamble	3.5
$= \dfrac{1}{\pi}\int_{-\infty}^{\infty} \dfrac{v(\lambda)}{t-\lambda}\,d\lambda$	Transformada de Hilbert	5.4

Funciones	
$Q(\kappa) = \dfrac{1}{\sqrt{2\pi}}\int_{\kappa}^{\infty} e^{-\lambda^2/2}\,d\lambda$	Probabilidad gaussiana
$\exp t = e^t$	Exponencial
$\text{senc}\, t = \dfrac{\text{sen}\,\pi t}{\pi t}$	Senc
$\text{sgn}\, t = \begin{cases} 1 & t>0 \\ -1 & t<0 \end{cases}$	Signo
$u(t) = \begin{cases} 1 & t>0 \\ 0 & t<0 \end{cases}$	Escalón
$\Pi\left(\dfrac{t}{\tau}\right) = \begin{cases} 1 & \|t\| < \dfrac{\tau}{2} \\ 0 & \|t\| > \dfrac{\tau}{2} \end{cases}$	Rectángulo
$\Lambda\left(\dfrac{t}{\tau}\right) = \begin{cases} 1 - \dfrac{\|t\|}{\tau} & \|t\| < \tau \\ 0 & \|t\| > \tau \end{cases}$	Triángulo

Símbolos misceláneos	
$\triangleq$	"Es igual por definición"
$\approx$	"Es igual en forma aproximada"
$\int_T$	$\int_{t_1}^{t_1+T}$ con t_1 es arbitrario
$\leftrightarrow$	Denota un par transformado de Fourier
$[a,b]$	El intervalo de a a b
★	Material del texto que se puede omitir

LECTURA SUPLEMENTARIA SELECCIONADA

continuación se listan bajo diferentes títulos, libros y documentos de utilidad para lectores que deseen un tratado alterno, o que deseen profundizar en algún tema. dos los artículos son susceptibles de encontrarse en una buena librería de libros nicos. A menos que se indique otra cosa, están a un nivel comparable con las secnes correspondientes de este texto. En la sección de referencias de dan citas mpletas.

temas de comunicación en general

s diferentes textos tienen más o menos la misma cobertura acerca de los sistemas comunicación que este libro. Los más recientes comprenden McMullen (1968), krison (1968), Schwartz (1970), Simpson y Houts (1971), y Taub y Schilling 71).

Goldman (1948), uno de los primeros libros sobre el tema, tiene muchos mplos de análisis espectral en el contexto de la ingeniería de radio. Bennett (1970) alca los aspectos de la comunicación más solucionados con los sistemas telefóni-, en tanto que Sunde (1969) trata en forma extensa los problemas en canales im-fectos. Los sistemas de comunicación por *láser* los cubre Pratt (1969).

Análisis de la señal

En Lathi (1965, Caps. 2 al 6), se encontrará un extenso análisis de señales y métodos de Fourier, así como en Frederick y Carlson (1971, Caps. 4 al 7 y 12). Dos libros que tratan en forma exclusiva las transformadas de Fourier y sus aplicaciones son Bracewell (1965) y Papoulis (1962). El primero se semeja a un diccionario gráfico de pares transformados; el último concierta un escrupuloso balance entre la exactitud y la claridad. En Lighthill (1958) y Franks (1969), que son trabajos más avanzados, son obras para el lector que desee profundizar.

Señales aleatorias

Beckmann (1967, Caps. 6 y 7), Lathi (1968, Caps. 2 al 4) y Thomas (1969. Caps. 2 al 4) dan exposiciones breves de teoría de probabilidades y de las señales aleatorias. De mayores alcances y más avanzados son Cooper y McGillem (1971), Davenport y Root (1958) y Papoulis (1965), en orden de mayor complejidad.

Han aparecido dos documentos excelentes en relación con el análisis del ruido: Rice (1944, 1945), una referencia clásica y Benett (1959), un artículo de orientación.

Ruido eléctrico

Baghdady (1960, Cap. 15) es un tratado conciso acerca del ruido en los receptores de comunicación. Una de las mejores referencias generales es Bennett (1960); además del ruido en el receptor, analiza el ruido en la radiación electromagnética, técnicas de medición y diseño de sistemas de bajo ruido.

El documento de Pierce (1956) examina las fuentes físicas del ruido. Siegman (1961) analiza el ruido en los sistemas de microondas, empleando la aproximación didáctica de la línea de transmisión. Van der Ziel (1970) trata de ruido en dispositivos de estado sólido y *láseres,* mientras que Jolly (1967) hace un examen cualitativo e ilustrativo de la electrónica de bajo ruido.

Modulación

Todos los textos en general, antes mencionados, dedican capítulos a la CW y a la modulación de pulsos. Además, Panter (1965) hace un extenso análisis de la FM y de la modulación por pulsos codificados y lo refuerza con abundantes citas, que lo hace un artículo valioso para la biblioteca de consulta de cualquier ingeniero en sistemas de comunicación.

En Taub y Schilling (1971, Caps. 9 y 10) se da una excelente presentación acerca del ruido y de la extensión del umbral de FM. Los dos documentos clásicos sobre la FM —Carlson (1922) y Armstrong (1936)— han resistido la prueba del tiempo y continúan siendo textos didácticos.

Aunque un poco cargado de datos, Black (1953, Caps. 4 y 5 y 15 al 20) cubre en forma excelente y total lo relativo al muestreo y la modulación de pulsos. Linden (1959) enfoca la teoría del muestreo; Rowe (1965, Cap. 4) cubre el análisis espectral

la modulación analógica de pulsos; y Oliver, Pierce y Shannon (1948) es el tículo insignia en la filosofía de la modulación por pulsos codificacos. Schindler 970) examina los diferentes tipos de sistemas de modulación de delta.

eoría de la comunicación estadística

as tres áreas principales de la teoría de la comunicación estadística (o sean, la teoría la información, la codificación y la teoría de la detección) son tratadas por Thoas (1969) y por Wozencraft y Jacobs (1965). El último requiere de mucha matemáca, pero bien vale la pena el esfuerzo.

Los documentos clásicos en la teoría de la información son Nyquist (1924, 28), Hartley (1928) y Shannon (1948, 1949). Shannon (1949) prescinde en forma lativa de las matemáticas y es altamente recomendable; entre otras cosas, este domento fue la primera exposición acerca de la teoría del muestreo aplicada a la counicación eléctrica. Abramson (1963), Ash (1965) y Gallager (1968) son textos a vel de graduados acerca de la teoría de la información y de la codificación. Por ra parte, Pierce (1961) fue escrito para el lego e incluye las implicaciones de la oría de la información en el arte, la música, la psicología, etc. A cualquier persona teresada en estos temas, le serán de utilidad las nuevas interpretaciones de Pierce.

La concisa monografía de Selin (1965) bosqueja la teoría de la detección. Hanock y Wintz (1966), y Brown y Palermo (1969), se encuentran a un nivel matemáti más elevado. Turin (1960) es un documento guía sobre filtros adaptados, ientras que el libro de Viterbi (1966) trata la detección coherente que emplea uipo coherente en fase.

omunicación digital de datos

ennett y Davey (1965), y Lucky, Salz y Weldon (1968) están dedicados por comple- a la transmisión digital de datos, siendo más teórico el último. Otros trabajos que ntienen una extensión apreciable sobre el tema comprenden a Schwartz, Bennett y ein (1966), Viterbi (1966), Wozencraft y Jacobs (1965).

De los muchos documentos que se podrían citar aquí, tienen un mérito especial s siguientes: Arthurs y Dym (1962), un análisis de la detección de señales digitales e descansa en el lenguaje del espacio de la señal; Salz (1965) describe la FM digital npleando discriminadores; Lucky (1965) escribió el artículo más importante sobre igualación de adaptación; y Jacobs (1967), compara los sistemas multiniveles.

Para una introducción a la codificación para control de error, Lin (1970) está tre los mejores. Berlekamp (1968), Peterson y Weldon (1972) dan presentaciones ás completas y avanzadas.

ircuitos de comunicación y la electrónica

veritt y Anner (1956), es uno de los textos modelos sobre líneas de transmisión, tros de ondas, etc.; Christian y Eisenmann (1966) , han compilado una tabulación conveniente de diseños de filtros. En Terman (1955) y Henney (1959) se encuentran muchos aspectos relacionados con los componentes físicos de la ingeniería de radio. Libros más recientes sobre la electrónica de la comunicación son DeFrance (1966), Zeines (1970), Clarke y Hess (1971), y Alley y Atwood (1973). Hansen (1969) está dedicado a la electrónica de la televisión.

EFERENCIAS

OTT, E.: "Flatland," 6th ed., rev., Blackwell, Oxford, 1950.

AMSON, N.: "Information Theory and Coding," McGraw-Hill, New York, 1963.

EY, C. L., and K. W. ATWOOD.: "Electronic Engineering," 3rd. ed., Wiley, New York, 1973.

ISTRONG, E. H.: A Method of Reducing Disturbances in Radio Signaling by a System of Frequency Modulation, *Proc. IRE*, vol. 24, pp. 689–740, May 1936.

HURS, E., and H. DYM: On the Optimum Detection of Digital Signals in the Presence of White Gaussian Noise—A Geometric Interpretation and a Study of Three Basic Data Transmission Systems, *IRE Trans. Commun. Systems*, vol. CS-10, pp. 336–372, December 1962.

, R.: "Information Theory," Wiley, New York, 1965.

HDADY, E. J. (ed.): "Lectures on Communication System Theory," McGraw-Hill, New York, 1960.

ER, P. A.: Phase-modulation Data Sets for Serial Transmission at 2000 or 2400 Bits per Second, *AIEE Trans.*, pt. I, vol. 61, pp. 161–171, July 1962.

KMANN, P.: "Probability in Communication Engineering," Harcourt, Brace & World, New York, 1967.

L TELEPHONE LABORATORIES: "Transmission Systems for Communication," revised 4th ed., Bell Telephone Laboratories, Winston-Salem, N.C., 1971.

BENNETT, W. R.: Methods of Solving Noise Problems, *Proc. IRE*, vol. 44, pp. 609–638, May 1956. "Electrical Noise," McGraw-Hill, New York, 1960. "Introduction to Signal Transmission," McGraw-Hill, New York, 1970.

BENNETT, W. R., and J. R. DAVEY: "Data Transmission," McGraw-Hill, New York, 1965.

BERLEKAMP, E. R.: "Algebraic Coding Theory," McGraw-Hill, New York, 1968.

BLACK, H. S.: "Modulation Theory," Van Nostrand, Princeton, N.J., 1953.

BRACEWELL, R.: "The Fourier Transform and Its Applications," McGraw-Hill, New York, 1965.

BRILLOUIN, L.: "Science and Information Theory," Academic, New York, 1956.

BROWN, W. M., and C. J. PALERMO: "Random Processes, Communications and Radar," McGraw-Hill, New York, 1969.

CARSON, J. R.: Notes on the Theory of Modulation, *Proc. IRE*, vol. 10, pp. 57–64, February 1922 (reprinted in *Proc. IEEE*, vol. 51, pp. 893–896, June 1963).

CHAFFEE, J. G.: The Application of Negative Feedback to Frequency-modulation Systems, *Proc. IRE*, vol. 27, pp. 317–331, May 1939.

CHRISTIAN, E., and E. EISENMANN: "Filter Design Tables and Graphs," Wiley, New York, 1966.

CLARKE, K. K., and D. T. HESS: "Communication Circuits: Analysis and Design," Addison-Wesley, Reading, Mass., 1971.

CLOSE, C. M.: "The Analysis of Linear Circuits," Harcourt, Brace & World, New York, 1966.

COOPER, G. R., and C. D. MCGILLEM: "Probabilistic Methods of Signal and System Analysis," Holt, New York, 1971.

CROISIER, A., and J. D. PIERRET: The Digital Echo Modulation, *IEEE Trans. Commun. Technology*, vol. COM-18, pp. 367–376, August 1970.

DAVENPORT, W. B., JR., and W. L. ROOT: "Introduction to Random Signals and Noise," McGraw-Hill, New York, 1958.

DEFRANCE, J. J.: "Communications Electronics Circuits," Holt, New York, 1966.

DOWNING, J. J.: "Modulation Systems and Noise," Prentice-Hall, Englewood Cliffs, N.J., 1964.

DRAKE, A. W.: "Fundamentals of Applied Probability Theory," McGraw-Hill, New York, 1967.

EVERITT, W. L., and G. E. ANNER: "Communication Engineering," 3rd ed., McGraw-Hill, New York, 1956.

FRANKLIN, R. H., and H. B. LAW: Trends in Digital Communication by Wire, *IEEE Spectrum*, vol. 3, pp. 52–58, November 1966.

FRANKS, L. E.: "Signal Theory," Prentice-Hall, Englewood Cliffs, N.J., 1969.

FREDERICK, D. K., and A. B. CARLSON: "Linear Systems in Communication and Control," Wiley, New York, 1971.

FRIEDMAN, B.: "Principles and Techniques of Applied Mathematics," Wiley, New York, 1956.

GALLAGER, R. G.: "Information Theory and Reliable Communication," Wiley, New York, 1968.

GOLDMAN, S.: "Frequency Analysis, Modulation, and Noise," McGraw-Hill, New York, 1948.

AMMING, R. W.: Error Detecting and Error Correcting Codes, *Bell System Tech. J.*, vol. 29, pp. 147–160, April 1950.

ANCOCK, J. C., and P. A. WINTZ: "Signal Detection Theory," McGraw-Hill, New York, 1966.

ANSEN, G. L.: "Introduction to Solid-State Television Systems," Prentice-Hall, Englewood Cliffs, N.J., 1969.

ARMUTH, H. F.: Applications of Walsh Functions in Communications, *IEEE Spectrum*, vol. 6, pp. 82–91, November 1969.

ARTLEY, R. V.: Transmission of Information, *Bell System Tech. J.*, vol. 7, pp. 535–563, July 1928.

ENNEY, K. (ed.): "Radio Engineering Handbook," 5th ed., McGraw-Hill, New York, 1959.

OROWITZ, M.: The Dolby Technique for Reducing Noise, *Popular Electronics*, pp. 31–34, August 1972.

COBS, I.: Comparison of *M*-ary Modulation Systems, *Bell System Tech. J.*, vol. 46, pp 843–863, May–June 1967.

HNKE, E., and F. EMDE: "Tables of Functions," 4th ed., Dover, New York, 1945.

HNSON, J. B.: Thermal Agitation of Electricity in Conductors, *Phys. Rev.*, vol. 32, pp. 97–109, July 1928.

LLY, W. P.: "Low Noise Electronics," American Elsevier, New York, 1967.

RGEN, R. K.: Untangling the "Quad" Confusion, *IEEE Spectrum*, vol. 9, pp. 55–62, July 1972.

THI, B. P.: "Signals, Systems and Communication," Wiley, New York, 1965. "An Introduction to Random Signals and Communication Theory," International Textbook, Scranton, Pa., 1968.

GHTHILL, M. J.: "An Introduction to Fourier Analysis and Generalized Functions," Cambridge, New York, 1958.

N, S.: "An Introduction to Error-Correcting Codes," Prentice-Hall, Englewood Cliffs, N.J., 1970.

IDEN, D. A.: A Discussion of Sampling Theorems, *Proc. IRE*, vol. 47, pp. 1219–1226, July 1959.

CKY, R. W.: Automatic Equalization for Digital Communication, *Bell System Tech. J.*, vol. 44, pp. 547–588, April 1965.

CKY, R. W., J. SALZ, and E. J. WELDON, JR.: "Principles of Data Communication," McGraw-Hill, New York, 1968.

MULLEN, C. W.: "Communication Theory Principles," Macmillan, New York, 1968.

HOLS, M. H., and L. L. RAUCH: "Radio Telemetry," 2nd ed., Wiley, New York, 1956.

QUIST, H.: Certain Factors Affecting Telegraph Speed, *Bell System Tech. J.*, vol. 3, pp. 324–346, April 1924. Certain Topics in Telegraph Transmission Theory, *Trans. AIEE*, vol. 47, pp. 617–644, April 1928. Thermal Agitation of Electric Charge in Conductors, *Phys. Rev.*, vol. 32, pp. 110–113, July 1928.

VER, B. M., J. R. PIERCE, and C. E. SHANNON: The Philosophy of PCM, *Proc. IRE*, vol. 36, pp. 1324–1332, November 1948.

NTER, P. F.: "Modulation, Noise, and Spectral Analysis," McGraw-Hill, New York, 1965.

PAPOULIS, A.: "The Fourier Integral and Its Applications," McGraw-Hill, New York, 1962. "Probability, Random Variables, and Stochastic Processes," McGraw-Hill, New York, 1965.

PETERSON, W. W., and E. J. WELDON, JR.: "Error Correcting Codes," 2nd ed., M.I.T. Press, Cambridge, Mass., 1972.

PIERCE, J. R.: Physical Sources of Noise, *Proc. IRE*, vol. 44, pp. 601–608, May 1956. "Symbols, Signals and Noise," Harper & Row, New York, 1961.

PRATT, W. K.: "Laser Communication Systems," Wiley, New York, 1969.

REEVES, A. H.: The Past, Present, and Future of PCM, *IEEE Spectrum*, vol. 2, pp. 58–63, May 1965.

RICE, S. O.: Mathematical Analysis of Random Noise, *Bell System Tech. J.*, vol. 23, pp. 282–332, 1944, and vol. 24, pp. 46–156, 1945. Statistical Properties of a Sine-wave plus Random Noise, *Bell System Tech. J.*, vol. 27, pp. 109–157, January 1948. Communication in the Presence of Noise—Probability of Error for Two Encoding Schemes, *Bell System Tech. J.*, vol. 29, pp. 60–93, January 1950.

ROWE, H. E.: "Signals and Noise in Communication Systems," Van Nostrand, Princeton, N.J., 1965.

SAKRISON, D. J.: "Communication Theory: Transmission of Waveforms and Digital Information," Wiley, New York, 1968.

SALZ, J.: Performance of Multilevel Narrow-band FM Digital Communication Systems, *IEEE Trans. Commun. Technology*, vol. COM-13, pp. 420–424, December 1965.

SANDERS, R. W.: Communication Efficiency Comparisons of Several Communication Systems, *Proc. IRE*, pp. 575–588, April 1960.

SCHINDLER, H. R.: Delta Modulation, *IEEE Spectrum*, vol. 7, pp. 69–78, October 1970.

SCHWARTZ, M.: "Information Transmission, Modulation, and Noise," 2d ed., McGraw-Hill, New York, 1970.

SCHWARTZ, M., W. R. BENNETT, and S. STEIN: "Communication Systems and Techniques," McGraw-Hill, New York, 1966.

SELIN, I.: "Detection Theory," Princeton University Press, Princeton, N.J., 1965.

SHANNON, C. E.: A Mathematical Theory of Communication, *Bell System Tech. J.*, vol. 27, pp. 379–423, July 1948, and vol. 27, pp. 623–656, October 1948. Communication in the Presence of Noise, *Proc. IRE*, vol. 37, pp. 10–21, January 1949.

SIEGMAN, A. E.: Thermal Noise in Microwave Systems, *Microwave J.*, vol. 4, pp. 81–90, March 1961, vol. 4, pp. 66–73, April 1961, and vol. 4, pp. 93–104, May 1961.

SIMPSON, R. S., and R. C. HOUTS: "Fundamentals of Analog and Digital Communication Systems," Allyn and Bacon, Boston, 1971.

SLEPIAN, D.: A Class of Binary Signaling Alphabets, *Bell System Tech. J.*, vol. 35, pp. 203–234, January 1956.

SOMMERVILLE, D. M.: "An Introduction to the Geometry of *N* Dimensions," Dutton, New York, 1929.

STILL, A.: "Communication through the Ages," Holt, New York, 1946.

STUMPERS, F. L.: Theory of Frequency-modulation Noise, *Proc. IRE*, vol. 36, pp. 1081–1902, September 1948.

SUNDE, E. D.: "Communication Systems Engineering Theory," Wiley, New York, 1969.

UB, H., and D. L. SCHILLING: "Principles of Communication Systems," McGraw-Hill, New York, 1971.
RMAN, F. E.: "Electronic and Radio Engineering," 4th ed., McGraw-Hill, New York, 1955.
OMAS, J. B.: "An Introduction to Statistical Communication Theory," Wiley, New York, 1969.
RIN, G. L.: An Introduction to Matched Filters, *IRE Trans. Inform. Theory*, vol. IT-6, pp. 311–329, June 1960.
N DER ZIEL, A.: Noise in Solid-state Devices and Lasers, *Proc. IEEE*, vol. 58, pp. 1178–1206, August 1970.
ERBI, A. J.: "Principles of Coherent Communication," McGraw-Hill, New York, 1966.
ENER, N.: "Extrapolation, Interpolation, and Smoothing of Stationary Time Series," Wiley, New York, 1949.
ZENCRAFT, J. M., and I. M. JACOBS: "Principles of Communication Engineering," Wiley, New York, 1965.
NES, B.: "Electronic Communication Systems," Prentice-Hall, Englewood Cliffs, N.J., 1970.

ÍNDICE ANALÍTICO

ISBN 968-6046-83-6